Large-Scale Loan Portfolio Selection

Justin A. Sirignano Gerry Tsoukalas Kay Giesecke*

May 28, 2016

Abstract

We consider the problem of optimally selecting a large portfolio of risky loans, such as mortgages, credit cards, auto loans, student loans, or business loans. Examples include loan portfolios held by financial institutions and fixed-income investors as well as pools of loans backing mortgage- and asset-backed securities. The size of these portfolios can range from the thousands to even hundreds of thousands. Optimal portfolio selection requires the solution of a high-dimensional nonlinear integer program and is extremely computationally challenging. For larger portfolios, this optimization problem is intractable. We propose an approximate optimization approach that yields an asymptotically optimal portfolio for a broad class of data-driven models of loan delinquency and prepayment. We prove that the asymptotically optimal portfolio converges to the optimal portfolio as the portfolio size grows large. Numerical case studies using actual loan data demonstrate its computational efficiency. The asymptotically optimal portfolio's computational cost does not increase with the size of the portfolio. It is typically many orders of magnitude faster than nonlinear integer program solvers while also being highly accurate even for moderate-sized portfolios.

*Sirignano (jasirign@stanford.edu) and Giesecke (giesecke@stanford.edu) are from Stanford University, Management Science and Engineering. Tsoukalas (gtsouk@wharton.upenn.edu) is from the Wharton School, University of Pennsylvania. The authors would like to thank Pierre Bonami (BONMIN), Nick Sahinidis (BARON) and Jonathan Currie (Opti Toolbox) for their help fine tuning the MINLP solvers used in the numerical studies section. We also thank participants at the IPAM Workshop on New Directions in Financial Mathematics for comments.

1 Introduction

Financial institutions such as banks and government sponsored enterprises such as Fannie Mae own as well as securitize portfolios of mortgages, credit card receivables, auto loans, student loans, and business loans. Various other entities, such as pension funds, mutual funds, hedge funds, insurance firms, and the US Treasury, also hold portfolios of loans or securities backed by loan portfolios. The size of the consumer and commercial loan market is larger than the corporate credit market and rivals the equity market. According to estimates by the Federal Reserve and the Securities Industry and Financial Markets Association, there was over \$13.4 trillion in mortgages, \$8.7 trillion in mortgage-related securities, \$3.3 trillion in consumer credit, and \$1 trillion in student loans outstanding at the end of 2014.

Optimal loan portfolio selection requires the solution of a high-dimensional nonlinear integer program (which is NP-hard) and is extremely computationally challenging. Portfolios can commonly have anywhere from hundreds to hundreds of thousands of loans (Melennec 2000*b*), and each loan is characterized by a high-dimensional vector of loan-level features such as credit score, interest rate, loan balance, collateral, purpose, payment history, and location. Optimal selection of a loan portfolio is an integer program since loans can only be held in unit amounts. Objective and constraint functions are nonlinear and sometimes nonconvex, and they can be computationally costly to evaluate. Securities backed by loan pools, such as collateralized loan and mortgage obligations, can be complex derivatives of the underlying loan portfolio.

This paper presents an approach for tractable large-scale optimization of loan portfolios. The perspective is that of a lender, investor, or asset-backed security structurer who seeks to select a portfolio of loans that optimizes a performance measure subject to a set of constraints. For example, an investor may want to choose a portfolio of loans that minimizes some risk measure subject to the expected return being greater than a chosen threshold. A structurer may be interested in selecting a pool of loans to back a collateralized loan obligation so as to minimize the risk to the senior tranche investor, subject to constraints on the composition of the pool. Our formulation of the objective and constraints is sufficiently general to encapsulate many practical examples, including lending portfolios, collateralized loan obligations, mortgage-backed securities, collateralized mortgage obligations, and other asset-backed securities.

We consider a broad class of dynamic loan-level models of prepayment and delinquency, which include generalized linear models and machine learning models such as logistic regression and neural networks. These "data-driven" models are widely used in practice and are fitted from historical loan performance data that are collected internally or acquired from data vendors. We harness the limiting laws for large pools of loans recently developed under these models by Sirignano & Giesecke (2014) to approximate the objective and constraint functions of the problem, forming an approximate optimization problem. We prove that as the size of the portfolio grows large, the solution to the approximate problem, which we refer to as the asymptotically optimal portfolio (AOP), converges to the true optimal portfolio (i.e., the solution of the actual integer optimization problem). The AOP has significant computational advantages. It is the solution to a continuous optimization problem (not an integer program) and its dimension is typically far smaller than the dimension of the exact integer program. The dimension of the exact integer program is the size of the pool of loans from which the portfolio is being selected. Thus, for portfolios being chosen from pools in the thousands, tens of thousands, or even hundreds of thousands, the integer program quickly becomes intractable. In contrast, the AOP's dimension does not increase with the size of the portfolio (or pool from which it is being selected) since it is solving for a distribution over the "loan types".

We extensively test our approach for a variety of problems, including several which use actual loan data. The numerical results highlight the computational efficiency of the AOP. Two numerical studies of note are the selection of prime and subprime mortgage portfolios under mean-variance and log-optimal objectives. In those two examples, we test the AOP using models that are fitted to historical mortgage default and prepayment data and pools of actual mortgages drawn from our data set. Several other numerical studies are also presented, including optimally selecting a portfolio to back an MBS and selecting a portfolio to maximize exponential utility. The AOP is compared against the true optimal portfolio, which is obtained by solving the actual nonlinear integer optimization problem using best-of-class nonlinear integer program solvers. The AOP consistently outperforms these nonlinear integer program solvers on every problem we study. The AOP is often many orders of magnitude faster than the integer program solvers. In fact, for larger portfolios (e.g., selecting several thousand loans from an available pool of tens of thousands of loans) the integer program solvers break down (ran out of memory) or take extremely long times to solve (days or even weeks). In contrast, the AOP is able to solve the problem in seconds. Moreover, for many of these problems, the integer program solvers' solutions are actually suboptimal compared to the AOP. That is, the exact objective function evaluated at the AOP is smaller than the exact objective function evaluated at the integer program solution (assuming one is minimizing the objective). For all the problems we study, the AOP agrees strongly with the best of the integer program solvers' solutions ($97 - 99$ percent agreement for the solution vectors).

1.1 Related Literature

Despite its obvious importance, the problem of selecting a portfolio of risky loans has not received nearly as much attention as the equity portfolio problem. Bennett (1984) provides an early discussion. Altman (1996) explores a standard mean-variance formulation of the problem, as well as an alternative formulation that uses the unexpected loss from defaults as a risk measure. Paris (2005) treats the selection problem for a portfolio of consumer loans using a one-period, discrete-state formulation and an expected utility objective. Mencia (2012) studies mean-variance and utility-based formulations when loans are placed into groups of loans with similar characteristics. None of these authors address the significant modeling and computational issues arising with large problems, nor the availability of detailed loan- and borrower-level information. The formulation of the loan portfolio problem we propose in this paper harnesses the historical loan performance data commonly available and facilitates the treatment of the large problems that are common in practice.

Andersson, Mausser, Rosen & Uryasev (2001), Akutsu, Kijima & Komoribayashi (2004), Kraft & Steffensen (2008), Meindl & Primbs (2006), Wise & Bhansali (2002), and others study static and dynamic corporate bond portfolio selection problems using expected utility or other objectives. Capponi & Figueroa-

Lopez (2014) and Capponi, Figueroa-Lopez & Pascucci (2015) develop regime-switching models to address the dynamic asset allocation problem in defaultable markets. Kraft & Steffensen (2009) study how contagion and bankruptcy procedures affect the selection problem. Giesecke, Kim, Kim & Tsoukalas (2014) analyze the static selection problem for a portfolio of credit swaps using a goal programming approach which includes real trading constraints. Bo & Capponi (2014) examine the dynamic selection problem for a portfolio of credit swaps using a power utility objective and dynamic programming approach. While bonds and credit swaps have features similar to those of loans, the problems studied in the aforementioned papers differ significantly from the class of problems we address in this paper. The bond and swap portfolio problems do not usually call for integer constraints. Prepayment risk is absent unless one considers callable bonds. Short positions can often be implemented, while it is rarely possible to short loans. Finally, the size of the bond and swap portfolio problems is typically much smaller, by orders of magnitude, than that of the loan portfolio problem.

Saunders, Xiouros & Zenios (2007) approximate a credit portfolio optimization problem by replacing the true optimization objective with a law of large numbers. Their model setting differs from ours in several ways. They consider a static model for the underlying assets while our model is dynamic. They consider an optimization problem where the goal is to select the portion of the portfolio in multiple discrete buckets (i.e., within each bucket the assets are homogeneous). One can choose to place a continuous amount of capital in each bucket; they therefore deal with a continuous optimization problem. We consider the problem of choosing a portfolio of heterogeneous loans where each loan's features take values in a continuous space. Therefore, one must optimize over an infinite-dimensional function (e.g., a measure on the real line). Loans in our setting can only be chosen in unit amounts, making the problem an integer program. Saunders et al. (2007) only include a constraint on short-selling, while we include general constraint functions. This, for instance, allows us to consider a mean-variance problem where the variance is minimized subject to the expected return being greater than some threshold. We also approximate the true portfolio optimization problem using both the law of large numbers and central limit theorem in order to increase the accuracy of our approximation. Finally, a major focus of our paper is computational methods for solving the approximate optimization problem for actual loan data and comparing the performance to integer program solvers.

Since the selection of a loan is generally a binary decision, our work can be related to the cardinality constrained portfolio optimization literature. This literature extends the standard mean-variance model for stocks to account for integer constraints, and develops relaxation methods which take advantage of the unique properties of the quadratic formulation. Among the first studies, Blog, Van der Hoeck, Rinnooy & Timmer (1983) propose a dynamic programming heuristic for small portfolios, and Bienstock (1996) proposes a surrogate constraint in lieu of the cardinality constraint. This approach is extended in Bertsimas & Shioda (2009), who propose a convex relaxation and pivoting method. More recently, Gao & Li (2013) develop a Lagrangian relaxation and geometric approach which exploits a special symmetric property of the quadratic objective. Our work significantly differs from these studies. First, we focus on data-driven models for loans, not stocks. The primary source of risk is event risk (delinquency, prepayment), rather than the risk of price movements. We face two sources of high-dimensionality: the large number of assets, and the large number of asset-level features. Second, our approach is based upon weak convergence results for large pools. It applies to many objectives and constraints, and does not rely on any restrictive or special (e.g., quadratic) structure of the problem.

1.2 Structure of Paper

Section 2 describes the general loan portfolio problem for a broad class of data-driven models of loan delinquency and prepayment. The approximation of the optimization problem and the asymptotically optimal portfolio are presented and analyzed in Section 3. Section 4 numerically implements our approach for several examples using actual loan data. Proofs of theorems can be found in an appendix.

2 Problem Formulation

2.1 Class of Models

For a broad family of discrete-time models for loan prepayment and delinquency, we consider the problem of optimally selecting N loans for an investment portfolio or a security backed by a portfolio of loans. We fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and an information filtration $(\mathcal{F}_t)_{t\geq 0}$, where $\mathbb{P}$ is the actual probability measure. The state of the n-th loan at time $t \in I = \{0, 1, \dots, T\}$ is $U_t^n \in \mathcal{U}$, where $\mathcal{U}$ is a finite discrete space, T is the time horizon, and $n \in \{1, \dots, N\}$. For example, if the loans are subject to both default and prepayment risk, $\mathcal{U} = \{\text{outstanding}, \text{prepaid}, \text{default}\}$. For notational convenience, we will refer to these states as o, p, and d, respectively, and we assume that p and d are absorbing states. Additional states could be 30 days late, 60 days late, etc. For each loan, $U_0^n = \text{o}$. Each loan has (static) features $Y^n \in \mathcal{Y} \subset \mathbb{R}^{d_Y}$. The feature space $\mathcal{Y}$ includes both continuous and categorical variables. For instance, Y^n might include features such as credit score, loan-to-value (LTV) ratio, initial interest rate, type of loan, collateral type, and geographic location. In addition, a stochastic process $X = (X_t)_{t\in I}$ with $X_t \in \mathbb{R}^{d_X}$ models common stochastic factors such as national interest rate and national unemployment rate that have an influence on all loans. The common factor X drives the correlation amongst the loans in the pool.

The conditional state transition probability takes the form

$$\mathbb{P}[U_t^n = u | \mathcal{F}_{t-1}] = h_\theta(u, U_{t-1}^n, Y^n, X_{t-1}), \qquad t \in \{1, \dots, T\}, \tag{1}$$

where the transition function h_θ is specified by a parameter $\theta \in \Theta$ that is estimated from loan performance data as in Khandani, Kim & Lo (2010), Sirignano, Sadhwani & Giesecke (2015), Banasik, Crook & Thomas (1999), Capozza, Kazarian & Thomson (1997), Stepanova & Thomas (2002), Baesens (2005), Bastos (2010), Westgaard & der Wijst (2001), and many others. A typical formulation for h_θ would be a generalized linear model (GLM), such as logistic regression, or a machine learning model, such as a neural network. To model seasonality of transitions, dependence of h_θ upon time can be incorporated by including time as an element of the common factor X.

Our discrete-time formulation is motivated by the data structures often encountered in practice. The loan performance data used for fitting the transition model (1) is usually updated monthly. However, all of the theoretical and computational results developed below can be extended to the case where the loan dynamics are modeled in continuous time. A continuous-time framework would employ stochastic intensities to model state transitions; a typical formulation might be a Cox proportional hazard model. The model can also be extended to include full path dependence upon X; see Sirignano & Giesecke (2014). In addition, one could allow the loan-level features Y^n to be independent stochastic processes (i.e., varying with time). Finally, one could include contagion effects via a mean-field term in (1). The results in the paper can be extended to these more general cases.

2.2 Loan Portfolio

A loan portfolio P^N is a selection of N loans from available loans in the universe $\mathcal{Y}$. In usual applications, N would be fixed in advance due to external regulatory, financial, and investor demand requirements on the size of a portfolio. Thus, a portfolio P^N is a choice of loans $\{Y^1, \ldots, Y^N\}$. Define

$$P^N = \frac{1}{N}\sum_{n=1}^{N} \delta_{Y^n},$$

where δ is the Dirac measure. This definition of P^N is equivalent to a vector of choices $(Y^1, \ldots, Y^N)$ for the N loans in the portfolio. The choice of a portfolio is integer-constrained since at a particular point $y \in \mathcal{Y}$, loans with the feature y can only be added to the portfolio in integer amounts. Since P^N takes values in the space of probability measures, only long positions are allowed (in practice, it is usually difficult to short a loan).[1] In addition, define the empirical measure $\mu_t^N \in \mathcal{M}(\mathcal{U} \times \mathcal{Y})$, where $\mathcal{M}(E)$ is the space of probability measures on E:

$$\mu_t^N = \frac{1}{N}\sum_{n=1}^{N} \delta_{(U_t^n, Y^n)}.$$

The empirical measure $\mu_t^N(u, dy)$ gives the fraction of loans occupying the state u with features y. More specifically, for a set $A \in \mathcal{Y}$ and a state u, $\mu_t^N(u, A)$ gives the fraction of loans in state u and with features in the set A. In addition, $\left\langle f, \mu_t^N(u, \cdot)\right\rangle_{\mathcal{Y}} \equiv \int_{\mathcal{Y}} f(y)\mu_t^N(u, dy) = \frac{1}{N}\sum_{n=1}^{N} f(Y^n)\mathbf{1}_{U_t^n = u}$ where we have defined $\langle f, \nu\rangle_E = \int_E f(y)\nu(dy)$. At the initial time, $\mu_0^N(\mathrm{o}, dy) = P^N(dy)$. The empirical measure μ_t^N completely encodes the relevant information for a pool of loans.[2] Using the empirical measure, a wide variety of performance measures for a pool of loans or a security backed by a portfolio of loans can be

[1] Although not explicitly explored in this paper, one could easily allow short positions by optimizing over the cross-product of two probability measures, $P_+ \times P_-$, where P_+ are the long positions and P_- are the short positions.

[2] We assume that there is no interest in permutations of loans with exactly the same features.

concisely formulated as:

$$R^N_{P^N} \;=\; f(\mu^N, X), \tag{2}$$

where $\mu^N = (\mu^N_t)_{t\in I}$, $f : B^{T+1} \mapsto \mathbb{R}^{d_R}$, and the space $B = \mathcal{M}(\mathcal{U} \times \mathcal{Y}) \times \mathbb{R}^{d_X}$. Note that the choice of the portfolio P^N completely determines the distribution of the random variable $R^N_{P^N}$. As illustrated in Section 2.3 below, the formulation (2) is sufficiently general to encapsulate many practical examples, including: loan portfolios, collateralized loan obligations (CLOs), passthrough mortgage-backed securities (MBS), collateralized mortgage obligations (CMOs), and other asset-backed securities (ABSs). Often, the performance measure $R^N_{P^N}$ will be the return for the security. The performance measure $R^N_{P^N}$ is allowed to be multi-dimensional in order to account for multiple performance criteria. For example, ABS are often tranched, with several classes of investors, each with their own return deriving from the cashflow from the underlying portfolio of loans. Finally, the common factor X is included as cashflows are typically discounted at some interest rate, which would be included as an element of X. We model the risk-free interest rate r_t for the time period $[t, t+1)$ by $r_t = r(X_t)$. The corresponding discount factor for a cashflow at time t is denoted as D_t and we denote $D(X) = (D_1, D_2, \ldots, D_T)$.

An investor is faced with the decision of how to optimally select N loans for a portfolio. The investor wishes to minimize some functional of the selected performance measure subject to constraints. We consider a static portfolio optimization problem where the investor selects their portfolio at time zero and holds it until the horizon T. The optimal portfolio $P^{N,*}$ solves the following problem:

$$\min_{P\in\mathcal{M}^N(\mathcal{Y})} V^N(P) \equiv v_2(\mathbb{E}[v_1(R^N_P)]) \quad \text{such that} \quad J^N(P) \equiv \mathbb{E}[g(R^N_P)] \geq c, \quad q(P) \leq d, \tag{3}$$

where the functions $q : \mathcal{M}(\mathcal{Y}) \mapsto \mathbb{R}^{d_Q}$, $g : \mathbb{R}^{d_R} \mapsto \mathbb{R}^{d_g}$, $v_1 : \mathbb{R}^{d_R} \mapsto \mathbb{R}^{d_v}$, and $v_2 : \mathbb{R}^{d_v} \mapsto \mathbb{R}$, and c, d are real-valued constants. We have also introduced the space

$$\mathcal{M}^N(\mathcal{Y}) = \left\{ \frac{1}{N} \sum_{n=1}^{N} \delta_{y^n} : y^1, \ldots, y^N \in \mathcal{Y} \right\} \subset \mathcal{M}(\mathcal{Y}).$$

The function $J^N(P)$ could for instance be the expected return. The function $V^N(P)$ could be the variance, probability that the return is below a certain threshold, the investor's expected utility, or higher-order moments of the distribution. More generally, V^N could even be the distance between the distribution of the return and some target distribution: simply define f in (2) to be a metric between R^N_P and the target distribution. It may be desirable to use more sophisticated risk measures than the variance because default and prepayment events generate highly non-normal distributions. The second constraint might model a requirement on the average credit score/rating (or any other characteristic) of the underlying loan portfolio. Such constraints are commonly required by the rating agencies in order to garner a certain rating, see Melennec (2000*a*) and Melennec (2000*b*). The second constraint could also be used to represent a limit in supply or regulatory requirements for certain types of loans in the portfolio. Typically, one will be selecting a portfolio from an available pool of loans. Thus, there may be a limit on how much of each type of loan in the space $\mathcal{Y}$ that one can select. The function $q(P)$ can be used to represent such a supply constraint. It can also be used to represent a constraint on the total notional of a portfolio if the feature space $\mathcal{Y}$ includes as one of its elements the size of the loan. All of these objectives and constraints are commonly encountered in practice.

2.3 Examples

We provide several examples of typical optimization problems. For the purpose of the examples, it is implicitly assumed that all cashflows are reinvested at the risk-free rate and that each loan has unit notional value. These restrictions are, however, not required under the general formulation.

Example 2.1 (*Loan Portfolio*)**.** Banks and investors often hold large portfolios of business or consumer loans. The loans composing these portfolios are typically subject to default and prepayment risk, and there is a tradeoff between higher interest rates for more risky loans and their increased default risk. The feature space $\mathcal{Y}$ might include the interest rate, geographic location, loan amount, loan term, etc. The common factor X might be the national unemployment rate. Let $\mathcal{U} = \{\mathrm{o}, \mathrm{p}, \mathrm{d}\}$. Assume that all payments are made monthly and each loan has the same maturity T. Each month, the payment amount is $a(t, Y^n)$. Given that the n-th loan prepays at time t, the amount prepaid is $c(t, Y^n)$. Let $\ell(Y^n, X_t)$ be the loss given default for the n-th loan provided that it defaults at time t. The investor's return from the portfolio is:

$$
\begin{aligned}
R^N_{P^N} \;&=\; \frac{1}{N}\sum_{n=1}^{N}\sum_{t=1}^{T} D_t\big[a(t, Y^n)\mathbf{1}_{U^n_t=\mathrm{o}} + c(t, Y^n)\big(\mathbf{1}_{U^n_t=\mathrm{p}} - \mathbf{1}_{U^n_{t-1}=\mathrm{p}}\big) \\
&+\; (1-\ell(Y^n, X_t))\big(\mathbf{1}_{U^n_t=\mathrm{d}} - \mathbf{1}_{U^n_{t-1}=\mathrm{d}}\big)\big].
\end{aligned}
$$

The portfolio return can be easily written in the form of (2):

$$
\begin{aligned}
R^N_{P^N} \;&=\; \sum_{t=1}^{T} D_t\big[\,\big\langle a(t,\cdot), \mu^N_t(\mathrm{o},\cdot)\big\rangle_{\mathcal{Y}} + \big\langle c(t,\cdot), \mu^N_t(\mathrm{p},\cdot) - \mu^N_{t-1}(\mathrm{p},\cdot)\big\rangle_{\mathcal{Y}} \\
&+\; \big\langle 1-\ell(\cdot, X_t), \mu^N_t(\mathrm{d},\cdot) - \mu^N_{t-1}(\mathrm{d},\cdot)\big\rangle_{\mathcal{Y}}\,\big].
\end{aligned}
$$

An optimization goal for such a portfolio would be to minimize the risk of the return subject to the expected return being above a certain threshold:

$$
\min_{P\in\mathcal{M}^N(\mathcal{Y})} V^N(P) \quad \text{such that} \quad \mathbb{E}[R^N_P] \geq c.
$$

The risk measure V^N might be variance or the probability that the return is less than a certain threshold. It could also be a linear combination of the central moments, accounting not just for the variance but also for higher order effects such as skewness and kurtosis.

Example 2.2 (*Collateralized Loan Obligation*)**.** A collateralized loan obligation (CLO) is a securitization of a large number of business loans. A CLO has a tranched structure; for this example, we examine a simple structure with two tranches (equity and senior). The underlying loans can have features Y^n such as credit ratings, type of loan, type of business, and interest rate. Loans with lower credit ratings will be more likely to default, but will also have higher interest rates. The coupon rate for the senior tranche is 4% and the senior tranche's attachment point is 20% . Any remaining interest from the underlying loans is paid to the equity tranche. In the event that a loan defaults, the recovery from that loan is entirely paid to the senior tranche. Other asset-backed securities have similar structures to CLOs and their optimization can be treated in a similar manner as in this example.

Let $\mathcal{U} = \{\mathrm{o}, \mathrm{d}\}$. Assume the loans have coupon payments $a(t, Y^n)$ and maturity T. At maturity T, the loans pay back their notional. In addition, all coupon payments are made monthly. As in Example 2.1, $\ell(Y^n, X_t)$ is the loss given default. $R^N_{P^N}$ will be a two-dimensional vector, with the first element being the return to the equity tranche and the second element being the return to the senior tranche:

$$
\begin{aligned}
(R^N_{P^N})_{\text{equity}} &= \sum_{t=1}^{T} D_t \max(\langle a(t,\cdot), \mu^N_t(\mathrm{o},\cdot)\rangle_{\mathcal{Y}} - \frac{4}{5} \cdot \frac{.04}{12}, 0) + D_T \max(\langle 1, \mu^N_T(\mathrm{o},\cdot)\rangle_{\mathcal{Y}} - \frac{4}{5}, 0), \\
(R^N_{P^N})_{\text{senior}} &= \sum_{t=1}^{T} D_t \min(\frac{4}{5} \cdot \frac{.04}{12}, \langle a(t,\cdot), \mu^N_t(\mathrm{o},\cdot)\rangle_{\mathcal{Y}}) + D_T \min(\langle 1, \mu^N_T(\mathrm{o},\cdot)\rangle_{\mathcal{Y}}, \frac{4}{5}) \\
&+ \sum_{t=1}^{T} D_t \langle 1 - \ell(\cdot, X_t), \mu^N_t(\mathrm{d},\cdot) - \mu^N_{t-1}(\mathrm{d},\cdot)\rangle_{\mathcal{Y}}.
\end{aligned}
$$

The senior tranche is sold to an investor, who desires to minimize their risk. The equity tranche is typically retained by the bank which structured the deal, and the bank might desire a minimum expected return on the equity tranche. There also may be additional constraints on the fraction of loans in certain rating categories and different industries, and on the average rating of the underlying loan portfolio. The optimization problem would then be:

$$
\min_{P \in \mathcal{M}^N(\mathcal{Y})} V^N(P) \quad \text{such that} \quad \mathbb{E}[(R^N_P)_{\text{equity}}] \geq c_1, \quad \mathbb{E}[(R^N_P)_{\text{senior}}] \geq c_2, \quad q(P) \leq d.
$$

V is a risk measure, which might for instance be the variance of the senior tranche return. More generally, V could even be the distance under some metric between the distribution of the senior tranche return and a target distribution. Other constraints of practical interest which can be easily included are requirements on the average rating, average coupon rate, average recovery rate of the underlying loan portfolio, and Moody's diversification ratio, which can be written as a continuous function of $(P^N, P^N \times P^N)$.

Example 2.3 (*Passthrough Agency Mortgage-backed Security)*. In a passthrough MBS, the cashflows from the underlying pool of mortgages are directly passed on to the MBS investor; see Fabozzi, Bhattacharya & Berliner (2010) for details and background. An agency MBS is guaranteed against loss from default by the government-sponsored enterprise (either Freddie Mac or Fannie Mae) which securitized the MBS. Moreover, the mortgages are high quality and have low default rates. Therefore, it is a common approach to model only prepayments in such pools.

An agency MBS investor is exposed to prepayment risk. If the mortgage rates fall, many mortgage holders will refinance and prepay their mortgages. When unemployment rises, refinancing is more difficult to obtain due to tighter credit conditions and prepayment rates fall. There are other loan-specific factors which may influence the propensity of a mortgage holder to prepay, including: credit score, debt-to-income ratio, property type, and the existence of a prepayment penalty in the mortgage contract. The common factors X might include the national mortgage rate and the national unemployment rate. The features $\mathcal{Y}$ would include the interest rate, credit score, prepayment penalty flag, occupancy status, debt-to-income ratio, loan-to-value ratio, property type, geographic location, and other factors.

The return for the passthrough security can be easily written in the form of (2). Let $\mathcal{U} = \{\mathrm{o}, \mathrm{p}\}$ and assume the mortgages are fully amortizing with monthly payments $a(t, y)$ and maturity T. Given that the n-th loan prepays at time t, the amount prepaid is $c(t, Y^n)$. We have

$$R^N_{P^N} \;=\; \sum_{t=1}^{T} D_t \big[\big\langle a(t,\cdot), \mu^N_t(\mathrm{o},\cdot) \big\rangle_{\mathcal{Y}} + \big\langle c(t,\cdot), \mu^N_t(\mathrm{p},\cdot) - \mu^N_{t-1}(\mathrm{p},\cdot) \big\rangle_{\mathcal{Y}} \big].$$

The optimization goal is to minimize variance, subject to the expected return being above a certain threshold:

$$\min_{P \in \mathcal{M}^N(\mathcal{Y})} \mathrm{Var}[R^N_P] \quad \text{such that} \quad \mathbb{E}[R^N_P] \geq c.$$

3 Asymptotically Optimal Portfolio

The optimization problem (3) is a nonlinear integer program. It is well-known that integer programs are NP-hard. Furthermore, the number of decision variables in (3) is equal to N, where N can be anywhere from thousands to even hundreds of thousands for typical portfolios encountered in practice. For large N, the objective function and constraints can be computationally costly to calculate, taking many hours or days for a single evaluation on a personal computer. Finally, the objective function may even be nonconvex. Due to all of these reasons, the optimization problem (3) is very computationally challenging to solve.

Instead of solving the computationally intensive problem (3), we propose to approximate R^N_P for large N using limiting laws for the pool of loans. Under this approximation, the optimization problem becomes one of choosing an optimal distribution of loans $P \in \mathcal{M}(\mathcal{Y})$, where $\mathcal{M}(E)$ is the space of probability measures on the space E. Therefore, the high-dimensional nonlinear integer program (3) is transformed into an optimization problem over a single function. The approximate optimization problem has significant computational advantages: it is a continuous optimization problem rather than an integer program. For large N, the computational cost of evaluating the approximate objective and constraint functions is many orders of magnitude less than the computational cost of evaluating the true objective and constraint functions in (3). See Section 3.5 for a detailed discussion of the computational advantages of the AOP over integer program solvers. As illustrated in Section 4, the solution to the approximate optimization problem, termed the "asymptotically optimal portfolio," is often accurate even for relatively small N (such as $N = 125$).

In this section, we first construct the approximate optimization problem. The solution of the approximate problem is then proven to converge asymptotically to the true optimal portfolio as $N \to \infty$. Finally, we present some explicit solutions for the asymptotically optimal portfolio in a special case.

3.1 Approximate Optimization

The *asymptotically optimal portfolio* $\bar{P}^{N,*}$ is the solution to the optimization problem:

$$\min_{P \in \mathcal{M}(\mathcal{Y})} v_2(\mathbb{E}[v_1(\bar{R}_P^N)]) \quad \text{such that} \quad \mathbb{E}[g(\bar{R}_P^N)] \geq c, \quad q(P) \leq d, \tag{4}$$

where $\bar{R}_P^N$ is the approximation for R_P^N. The approximation is based upon a law of large numbers and central limit theorem for a pool of N loans. If the asymptotically optimal portfolio is solely based upon the law of large numbers, $\bar{R}_P^\infty$, it is denoted as $\bar{P}^{\infty,*}$.

For the class of models (1), Sirignano & Giesecke (2014) prove a dynamic law of large numbers and a dynamic central limit theorem. They assume the states $U_t^1, \ldots, U_t^N$ to be independent conditional upon $\mathcal{F}_{t-1}$ (weakening of this condition may be possible, but was not explored in that paper). The empirical measure μ^N converges in distribution to a limiting measure $\bar{\mu}$ in $\mathcal{M}(\mathcal{U} \times \mathcal{Y})^{T+1}$ as $N \to \infty$. The limiting measure $\bar{\mu}_t$ satisfies a dynamic, random equation, driven by the common factors X. The empirical fluctuation measure $\bar{\Xi}^N$ converges in distribution to the limiting distribution $\bar{\Xi}$ in W^{T+1} as $N \to \infty$, where $\Xi_t^N = \sqrt{N}(\mu_t^N - \bar{\mu}_t)$, $W = \prod_{u=1}^{|\mathcal{U}|} S'(\mathbb{R}^{d_Y})$, and S' is the space of tempered distributions. The variable $\bar{\Xi}_t$ is conditionally Gaussian given a path of the common factors X up to t. The law of large numbers satisfies:

$$\bar{\mu}_t(u, dy) = \sum_{u' \in \mathcal{U}} h_\theta(u, u', y, X_{t-1}) \bar{\mu}_{t-1}(u', dy). \tag{5}$$

It is important to note that the law of large numbers is dynamic and is also a random equation; randomness enters through the factor X. The law of large numbers has a natural link with the original model (1): the transition function h_θ appears in (5). The central limit theorem satisfies:

$$\bar{\Xi}_t(u, dy) = \sum_{u' \in \mathcal{U}} h_\theta(u, u', y, X_{t-1}) \bar{\Xi}_{t-1}(u', dy) + M_t(u, dy). \tag{6}$$

Given X, $M(u, dy)$ is a Gaussian process with zero mean and covariance satisfying

$$\begin{aligned}
\text{Cov}\Big[M_t(u_1, dy), M_t(u_2, dy) \big| X_{0:t-1}\Big] &= -\sum_{u' \in \mathcal{U}} h_\theta(u_1, u', y, X_{t-1}) h_\theta(u_2, u', y, X_{t-1}) \bar{\mu}_{t-1}(u', dy), \\
\text{Var}\Big[M_t(u, dy) \big| X_{0:t-1}\Big] &= \sum_{u' \in \mathcal{U}} h_\theta(u, u', y, X_{t-1})(1 - h_\theta(u, u', y, X_{t-1}) \bar{\mu}_{t-1}(u', dy),
\end{aligned}$$

where $u_1 \neq u_2$ and $X_{0:t} = (X_0, \ldots, X_t)$. Like the law of large numbers, the central limit theorem is also dynamic. Randomness for the limiting process $\bar{\Xi}$ enters both through X and a martingale term M. More details on the limiting equations $\bar{\mu}$ and $\bar{\Xi}$ can be found in Appendix B.

One can approximate a finite pool of size N using the law of large numbers and the central limit theorem. For large N, we have the following approximation:[3]

$$\mu_t^N \overset{d}{\approx} \bar{\mu}_t^N = \bar{\mu}_t + \frac{1}{\sqrt{N}} \bar{\Xi}_t. \tag{7}$$

The large pool approximation (7) is conditionally Gaussian and can be easily simulated. Conditional on each path of the common factors X, the distribution of $\bar{\mu}_t^N$ is Gaussian; the conditional mean and conditional covariance can be computed in closed-form. For large pools, the computational cost of solving (7) is typically several orders of magnitude less than the computational cost of brute-force simulation of (1). Moreover, even for pools with only a few hundred loans, the approximation is highly accurate. See Sirignano & Giesecke (2014) for a detailed description of computational methods for the numerical solution of (7). The large pool approximation (7) can be used to approximate the portfolio return R_P^N by simply setting the initial conditions $\bar{\mu}_0(\mathrm{o}, dy) = P(dy)$ and $\bar{\Xi}_0 = 0$. We make this explicit in the following notation. Denote $\bar{\mu}^\nu$ as the law of large numbers with initial condition $\bar{\mu}_0(\mathrm{o}, dy) = \nu(dy)$ (and zero for all other states $u \neq \mathrm{o}$). Also, let $\bar{\mu}^{N,\nu}$ be the large pool approximation with initial conditions $\bar{\mu}_0(\mathrm{o}, dy) = \nu(dy)$ (and zero for all other states $u \neq \mathrm{o}$) and $\bar{\Xi}_0 = 0$. The approximation for (2) is:

$$R_P^N \overset{d}{\approx} \bar{R}_P^N = f(\bar{\mu}^{N,P}, X), \tag{8}$$

where $\bar{\mu}^{N,P} = (\bar{\mu}_t^{N,P})_{t \in I}$. That is, we have chosen the initial composition of the portfolio according to the measure $P(dy)$, which is the quantity we will optimize over.

3.2 Convergence Analysis

In a previous paper, Sirignano & Giesecke (2014), the authors showed that the large pool approximation $\bar{\mu}^N$ will be close to the empirical measure μ^N for sufficiently large N. However, the convergence of the finite optimal portfolio to the limiting optimal portfolio is not automatically implied by $\mu^N \overset{d}{\to} \bar{\mu}^\infty$ and is much more challenging to prove.[4]

---

[3]The notation $\overset{d}{\approx}$ means the approximation holds in distribution. Rigorously, this means that $\rho_{1,d_R}(\nu^N, \bar{\nu}^N) \to 0$ as $N \to \infty$ where ν^N is the measure of the random variable $R_P^N \in \mathbb{R}^{d_R}$, $\bar{\nu}^N$ is the measure of $\bar{R}_P^N \in \mathbb{R}^{d_R}$, and $\rho_{1,d}$ is the Prokhorov metric for the space of probability measures on $\mathbb{R}^d$. Since $\bar{\mu}^N$ is distribution-valued, convergence for (7) is defined in terms of test functions. For any bounded function $\phi(u, y) : \mathcal{U} \times \mathcal{Y} \to \mathbb{R}$ which is continuous in y, $\rho_1(\nu_\phi^N, \bar{\nu}_\phi^N) \to 0$ as $N \to \infty$. Here, $\bar{\nu}_\phi^N$ is the measure of the real-valued random variable $\sum_{t \in I, u \in \mathcal{U}} \langle \phi(u, y), \bar{\mu}_t^N(u, dy) \rangle_{\mathcal{Y}}$, and ν_ϕ^N is the measure of the real-valued random variable $\sum_{t \in I, u \in \mathcal{U}} \langle \phi(u, y), \mu_t^N(u, dy) \rangle_{\mathcal{Y}}$.

[4]The finite optimization problem considered in this paper (without constraints) can be formulated as maximizing a function $G^N(z^N) : \mathcal{Z}^N \to \mathbb{R}$ where $\mathcal{Z}^N \subseteq \mathcal{Z}$. In the context of this paper, $G^N(z^N)$ could for instance be $\mathbb{E}[v_1(R_{z^N}^N)]$ where $\mathcal{Z}^N = \mathcal{M}^N(\mathcal{Y})$ and z^N is the selected portfolio. By the weak convergence results, we have that $G^N(z^N) \to G(z)$ as $N \to \infty$ if $z^N \to z \in \mathcal{M}$ as $N \to \infty$. However, this is in general not sufficient to show that $Z^{N,*} = \arg\max_{z^N \in \mathcal{Z}^N} G^N(z^N)$ converges to $Z^* = \arg\max_{z \in \mathcal{Z}} G(z)$ since arg max *is not a continuous operator*. We provide a simple counterexample, taken from van Handel (2008), where $\mathcal{Z}^N = \mathcal{Z} = [-1, 1]$, $G^N(z) = e^{-z^2} + 2e^{-(Nz - N + \sqrt{N})^2}$, and $G(z) = e^{-z^2}$. Even though $G^N(z) \to G(z)$ for every $z \in [-1, 1]$ and $\mathcal{Z}$ is compact, $Z^{N,*}$ does not converge to Z^*. Instead, $Z^{N,*} \to 1$ while $Z^* = 0$. In general, $G^N(z) \to G(z)$ for every $z \in \mathcal{Z}$ does not mean that $\arg\max_{z \in \mathcal{Z}} G^N(z)$ converges to $\arg\max_{z \in \mathcal{Z}} G(z)$. Proving the convergence of the optimums is often a very challenging problem.

In this paper, we show that the finite optimal portfolio $P^{N,*}$ weakly converges to the limiting optimal portfolio $\bar{P}^{\infty,*}$ (the "asymptotically optimal portfolio"). Theorem 3.1 below provides a theoretical guarantee for the AOP; namely, that for sufficiently large N, the AOP will be accurate, and its accuracy will increase as the portfolio size grows.

Theorem 3.1. *Suppose that $\mathcal{Y}$ is compact, the functions h_θ, g, v_1, v_2, and q are continuous, and there is a unique minimizer $P^{\infty,*}$ of (4) for $N = \infty$. Furthermore, suppose $f : B^{T+1} \times \mathbb{R}^{d_X} \to \mathbb{R}^{d_R}$ is of the form $f(\mu, x) = F(\mu, D(x))$ where $D : \mathbb{R}^{d_R} \to \mathbb{R}^{d_D}$ is a continuous bounded function and $F : B^{T+1} \times \mathbb{R}^{d_D} \to \mathbb{R}^{d_R}$ is continuous on $B^{T+1} \times \mathbb{R}^{d_D}$.[5] Finally, assume the optimization problem (3) is feasible for all N. Then, the optimal portfolio $P^{N,*}$ weakly converges to the limiting portfolio $\bar{P}^{\infty,*}$ as $N \to \infty$.*

The proof can be found in the Appendix A. As a corollary, $\rho(\bar{P}^{N,*}, P^{N,*}) \to 0$ as $N \to \infty$ where ρ is the Prokhorov metric (also see Appendix A). From a practical perspective, Theorem 3.1 shows that the limiting portfolio $\bar{P}^{\infty,*}$ is accurate for large N. Similarly, the asymptotically optimal portfolio $\bar{P}^{N,*}$ is also accurate for large N. In practice, the approximation $\bar{P}^{\infty,*}$ (based on the law of large numbers) will be sufficient for larger portfolios, although the approximation $\bar{P}^{N,*}$ (which includes both the law of large numbers and central limit theorem) will add some accuracy for moderate-sized portfolios.

The assumptions in Theorem 3.1 are mild and realistic to the loan portfolio setting. The strongest condition, that $P^{\infty,*}$ is the unique maximizer of (4) when $N = \infty$, is a standard assumption in the optimization and statistics literature. If the limiting optimization problem is convex, it will automatically be satisfied. The restriction that D is bounded is not limiting since in typical applications $D(x)$ will be a discount factor for the cashflows. The majority of the standard machine learning and statistical models used for the transition function h_θ in equation (1) will be continuous; examples include generalized linear models, neural networks, and Gaussian process regression.

Finally, one might wonder why we have restricted ourselves to the constraint $q(P) \leq d$ where q must be continuous in Theorem 3.1. A more general constraint might be $P(B) \leq \nu(B), \ \forall B \in \mathcal{Y}_B$ where ν is some measure and $\mathcal{Y}_B$ is a collection of closed sets in $\mathcal{Y}$. The obstacle to such a general formulation is that the set $\{P(B) \leq \nu(B), \ B \in \mathcal{Y}_B\}$ is not compact, and therefore the optimization problem (4) may not have a solution. However, for almost all practical applications, it should be noted that the constraint $q(P) \leq d$ is sufficient. To solve for the $\bar{P}^{\infty,*}$, one must typically discretize the space $\mathcal{Y}$ into a grid or "computational cells". Then, for computational purposes, the optimization is over a discrete space (i.e., $\mathcal{Y}$ would be treated as a finite discrete space), and hence q will always be continuous. Another approach would be to approximate constraints of the form $P(B) \leq \nu(B), \ \forall B \in \mathcal{Y}_B$ by approximating an indicator function using bump functions. Indicator functions can be arbitrarily closely approximated with a smooth bump function (for instance, a sigmoidal function). Therefore, constraints on closed sets of the real line can be closely approximated using such analytic functions.

[5] B and $\mathbb{R}^{d_X}$ are metric spaces under the Prokhorov and Euclidean metrics, respectively. $B^{T+1} \times \mathbb{R}^{d_D}$ is also a metric space under the appropriate product metric, and continuity of F is defined with respect to this product metric.

3.3 Some explicit solutions

Generally, there are no explicit solutions for the asymptotically optimal portfolio and (4) must be numerically solved. However, for certain simple cases, there are explicit solutions using calculus of variations. Define the mean and covariance functions $m(y)$ and $\sigma(y, y')$:

$$\begin{aligned} m(y) &= \mathbb{E}[f(\bar{\mu}^{N,\Psi_y}, X)], \\ \sigma(y, y') &= \text{Cov}[f(\bar{\mu}^{N,\Psi_y}, X), f(\bar{\mu}^{N,\Psi_{y'}}, X)], \end{aligned}$$

where $\Psi_y = \delta_y$. Suppose that f is linear in $\bar{\mu}_t^N$ and $\mathcal{Y} = [a_1, a_2] \subset \mathbb{R}$. Let p be the density of the measure P. Then,

$$\min_{p \in C(\mathcal{Y}), P(dy) = p(y)dy} \gamma \text{Var}[\bar{R}_P^N] - \mathbb{E}[\bar{R}_P^N] = \gamma \int_{a_1}^{a_2} \int_{a_1}^{a_2} \sigma(y, y') p(y) p(y') dy' dy - \int_{a_1}^{a_2} m(y) p(y) dy,$$
$$\int_{a_1}^{a_2} p(y) dy = 1, \tag{9}$$

where $C(\mathcal{Y})$ is the set of continuous functions on $\mathcal{Y}$. If $\sigma(y, y') = 0$ for $y \neq y'$, calculus of variations yields:

$$p(y) = \frac{m(y)}{2\gamma\sigma(y, y)} + \frac{c}{2\gamma\sigma(y, y)}, \tag{10}$$

where c is a constant and can be determined from the condition $\int_{a_1}^{a_2} p(y)dy = 1$. Importantly, note that (10) is only the solution if $p(y) \geq 0$ for all $y \in \mathcal{Y}$; otherwise, one must allow short-selling of loans for (10) to be the correct solution (see Footnote 1 for how to extend the original framework to allow for short-selling). If short-selling were allowed, $P(dy) = p(dy)dy$ would be a signed measure. The solution (10) looks similar to the Sharpe ratio. If $\sigma(y, y') \neq 0$ for $y \neq y'$, one can again use calculus of variations to obtain the Fredholm integral equation of the first kind:

$$m'(y') = 2\gamma \int_{a_1}^{a_2} \frac{\partial}{\partial y'} \sigma(y, y') p(y) dy, \tag{11}$$

which can be solved to find $p(y)$. The Fredholm integral equation has been extensively studied and is known to have a solution in the form of the Liouville-Neumann series. As before, (11) is only the correct solution if $p(y) \geq 0$ (otherwise short-selling must be allowed).

Although the cases with explicit solutions are limited in scope, they offer some insight into the solution of the asymptotically optimal portfolio. Without the systematic risk (no covariance between loans with features y and y' where $y \neq y'$), the formula (10) looks similar to the Sharpe ratio. In particular, there is a clear tradeoff in the formula between the idiosyncratic risk and the return of a loan y. Although the solution $p(y)$ does depend upon the solution $p(y')$ (through the constant c), such dependence could be said to be weak. In the case of systematic risk, the solution at $p(y)$ in (11) strongly depends upon the global solution through an integral equation.

3.4 Computational Approaches to Solving the AOP

The AOP's solution is a single function P, which is a measure on $\mathcal{Y}$. In practice, one cannot directly solve for the function P but must instead either discretize $\mathcal{Y}$ into grid points or use a set of basis functions. In the numerical studies in Section 4 below, we use the former approach. Various methods can be used to choose grid points. A simple approach which works well for small pools is assigning the grid points to be the values for the loans in the pool from which the portfolio is being chosen. However, when $\mathcal{Y}$ is high-dimensional and the size of the pool is large, a sparse grid is necessary. We use k-means clustering in order to choose the sparse grid. Relatively few grid points are typically needed in order to obtain an accurate solution; we explain the reasons for this in Section 3.5.

There are also other large-scale optimization methods that could potentially be applied to solving the AOP and might offer promising directions for future research. High-dimensional optimization problems have been previously considered, especially in the context of dynamic programming problems. Several approaches have been developed and successfully implemented. Tsitsiklis & Van Roy (2001), Longstaff & Schwartz (2001), and Bellman & Dreyfus (1959) use basis functions to solve high-dimensional dynamic programming problems. The first two papers especially focus on the problem of high-dimensional American options. Doya (2000) also uses basis functions for optimization, but in the context of reinforcement learning. Bokanowski, Garcke, Griebel & Klompmaker (2013) and Munos & Moore (2002) investigate sparse grids for optimization. Lewis & Nash (2005), Borzi & Schulz (2009), and Dreyer, Maar & Schulz (2000) develop multigrid methods for optimization.

3.5 Sources of Computational Advantages of AOP

There are several sources for the AOP's lower computational cost in comparison to solving the original integer program. Firstly, integer programs are much more difficult, and computationally time-consuming, than continuous optimization problems. Integer programs are NP-hard. The computational complexity of an integer program grows exponentially with the number of decision variables (Li & Sun 2006). This begins to pose severe challenges for high-dimensional optimization problems (such as loan portfolio selection), which will be discussed in more detail below. Moreover, for portfolio optimization, one has a *nonlinear* integer program, which makes the problem even more challenging (Hemmecke, Koppe, Lee & Weismantel 2010). Finally, many loan portfolio optimization problems are nonconvex; one example is optimizing over the tranches of an asset-backed security. These integer programs are even more difficult (Burer & Letchford 2012). In contrast, the AOP is able to take advantage of the insight that the finite optimal portfolio is close to the solution of a continuous optimization problem. Instead of solving a challenging nonlinear integer program for the finite optimal portfolio, the AOP solves a continuous optimization problem which is considerably less computationally expensive.

Secondly, the computational cost of evaluating the objective and constraint functions for the AOP is much lower than for the integer program. We give a rough estimate below. Suppose there are N loans. To calculate the objective and constraint functions, one must simulate the dynamics or evaluate outcome of each of the N loans. If N is large, this will take a long time. The AOP is numerically solved by discretizing $\mathcal{Y}$ into N_g grid points. At each grid point, the LLN and CLT must be simulated. The cost for simulating a single loan and the limiting laws at a single point are roughly the same. Therefore, the ratio of computational times for a single evaluation of the objective and constraint functions is:

$$\frac{\text{Computational time for a single evaluation of the objective/constraints for AOP}}{\text{Computational time for a single evaluation of the objective/constraints for integer program}} = \frac{N_g}{N}.$$

Typically, very few points N_g are necessary to accurately evaluate the LLN and CLT (see Section 4 of this paper and Sirignano & Giesecke (2014)). Furthermore, the number of grid points N_g remains constant no matter the size of the finite portfolio N. Later in this section, we provide more discussion on why so few grid points are needed for the AOP. For large N (for example, tens of thousands or hundreds of thousands of loans), the cost to evaluate the objective and constraint functions will be much lower for the AOP. Similarly, the ratio of the costs for a single evaluation of the gradients for the respective problems will be $\frac{N_g}{N}$. Additionally, in some cases, closed-form solutions are available for the AOP while they are not available for the integer program. This further increases cost savings for the AOP.

Thirdly, the integer program typically has a much higher dimension (i.e., many more optimization variables) than the AOP. The number of decision variables for the original integer program is N, the size of the portfolio. In contrast, the AOP's solution is a single function and can typically be numerically solved at a dimension much smaller than N. As previously mentioned, the computational complexity for an integer program grows exponentially with the number of decision variables. In our numerical studies reported in Section 4, the well-known solver BARON even has difficulties for portfolios where N is in the hundreds. There are additional computational challenges for high-dimensional optimization including memory constraints. Even if the problem were continuous, high-dimensions can pose large challenges. Second-order solvers (such as interior-point algorithms, trust region methods, and Newton methods) require solving linear systems of equations and will have complexity $O(N^3)$. This quickly becomes computationally infeasible for large N. One can still use first-order methods such as gradient descent, but the convergence rate is very slow and gradient descent is known to have difficulties with saddle points and local minima. Furthermore, for large N, the memory required to store the objective or constraint functions may become infeasible. For a mean-variance problem, the covariance matrix will have N^2 elements. A standard desktop computer will run out of memory for N in the tens of thousands. For all of these reasons, nonlinear integer program solvers will become computationally or even infeasible for larger portfolios.

The advantage of the AOP is that optimization occurs over a single function P instead of N loans where N can be large. Optimizing over a function allows one to take advantage of the inherent structure of the problem. Since the objectives and contraints are continuous functions on $\mathcal{Y}$, one expects the solution at points y_1 and y_2 to be very similar if y_1 and y_2 are sufficiently close in distance. For instance, if loans with feature y_1 are attractive to hold in the portfolio, it is very likely that loans with feature y_2 are also attractive and will be held in the portfolio. Thus, by optimizing over a single function, one is able to take advantage of this structure and achieve some *generalization*. One would expect the function to have some smoothness and continuity and moreover, within an appropriately small region in space, to not change too rapidly. Importantly, functions with suitable regularity can be arbitrarily well-approximated by other functions (for instance, basis functions or, more directly, by grid discretization). This is the principle that underpins the efficient solution of many infinite-dimensional problems, such as PDEs, machine learning (neural networks, Q-learning, etc.), and HJB equations. Similar to these other applications, the number of functions (or grid points) needed to approximate P tends to not be large when P has some structure to it; moreover, this number remains constant in N. If one treates every loan as a separate variable to optimize over (i.e., the number of decision variables is N), this structure is completely ignored and *all generalization is lost*.

Of course, for numerical implementation, one cannot directly solve for the function P but must instead either discretize $\mathcal{Y}$ into grid points or use a set of basis functions. However, the spirit of the argument carries over: due to the structure of the problem, relatively few grid points (i.e., the grid does not need to be very fine) or basis functions may be needed. Moreover, the number of grid points N_g does not increase with N but remains constant as N grows. Thus, the dimension of the AOP remains constant while the dimension N of the integer program becomes large. For example, if the dynamics of loans with features y_1 and y_2 are very similar and $||y_1 - y_2||$ is small enough, it is not necessary to have two grid points at y_1 and y_2, respectively, but instead suffices to have a single grid point $\frac{y_1+y_2}{2}$. The AOP takes advantage of the fact that by leveraging the LLN and CLT, one can group similar loan types together. Consequently, relatively large mesh sizes can be taken for the grid, resulting in relatively few decision variables for the AOP.

Finally, we again highlight the number of decision variables for the AOP (after discretization or basis functions) remains constant no matter the size of the original portfolio N. Consequently, the AOP does not have the drawback of large memory consumption and one can employ second-order solvers with fast convergence properties.

As will be seen in Section 4, the AOP is highly accurate and has a much lower computational cost than integer program solvers. Its accuracy can be directly attributed to the convergence of the finite optimal portfolio to the AOP as N grows. This convergence is theoretically guaranteed by Theorem 3.1. Intuitively, the fast convergence is due to the idiosyncratic noise quickly averaging out for the system, which can be seen in the numerical studies in Sirignano & Giesecke (2014).

4 Numerical Studies

To demonstrate the advantages of our optimization approach, we now numerically implement the asymptotically optimal portfolio (AOP) for several loan portfolio selection problems. The approximate optimization problem is solved using Matlab's basic interior point algorithm. We compare the AOP to solutions from various popular mixed integer nonlinear program (MINLP) solvers for the true optimization problem. These solvers include BARON and BONMIN. See Tawarmalani & Sahinidis (2005) and Bonami, Biegler, Conn, Cornuejols, Grossmann, Laird, Lee, Lodi, Margot & Waechter (2008), respectively, for descriptions of the solvers. BARON is generally considered the state of the art in terms of global mixed integer nonlinear program solvers. A detailed comparison of BARON against a range of other solvers is available in Neumaier, Shcherbina, Huyer & Vinko (2005). BARON requires a commercial license and can utilize IBM's CPLEX solver, which also requires a commercial license. BONMIN is also considered the state of the art amongst the freely available integer programming solvers. BONMIN makes use of several heuristics, most prominently its diving heuristic (see Bonami & Goncalves (2012)). These heuristics can be very successful in practice, but in general are not guaranteed to converge, whereas the branch and bound method has theoretical guarantees for convergence. In general, we find that BARON outperforms BONMIN when N is very small, but BONMIN (with or without heuristics) outperforms BARON when N becomes even moderately large. BONMIN's diving heuristic outperforms BONMIN's branch and bound for larger problems.

We consider a variety of test problems, which are described in more detail later in this section. First, we compare the performance of the AOP with the integer program solvers for a one-period model, which is a special case of our model class (1) with $T = 1$. In many cases, a one-period model allows for closed-form objective and gradients for both the AOP and the integer program. In addition, we compare the performance of the AOP with the integer program solvers for the full multi-period model ($T > 1$). In this latter case, the objective and gradients are approximated using Monte Carlo simulation.

(i) One-period Model ($T = 1$):

  (a) Selecting a mean-variance portfolio of loans

  (b) Selecting an optimal portfolio of loans under exponential utility

  (c) Selecting an optimal portfolio backing an equity tranche of a MBS

  (d) Selecting a geographically diverse mean-variance portfolio

(ii) Multi-period Model ($T > 1$):

  (a) Selecting a mean-variance portfolio of prime mortgages

  (b) Selecting a log-optimal portfolio of subprime mortgages

In each test case, we consider the optimization problem of selecting N loans from a pool of N_p available loans. Such a problem might be faced by a lender, investor, or asset-backed security structurer. A moderate-sized problem would have $N = 250$ and $N_p = 1,000$. Larger problems in practice could have $N = 2,500$ and $N_p = 10,000$ or $N = 25,000$ and $N_p = 100,000$. For the one-period model, the N_p available loans are randomly generated. For the multi-period model, the pool of N_p available loans are drawn from actual loan data sets. Default and prepayment model parameters for the multi-period model are fitted to actual loan data.

In all cases, we find that the AOP strongly outperforms BARON and BONMIN. The AOP has a much lower computational cost than the MINLP solvers and has a similar level of accuracy. In some cases, the AOP solution is actually better than the MINLP solvers' solutions. This means that the true objective function (V^N from the true optimization problem (3)) evaluated at the AOP solution is smaller than the true objective function evaluated at the MINLP solver's solution. In the remainder of the cases, the AOP solution is consistently close to the best solution amongst the MINLP solvers. For moderately large N (e.g., $N = 250$ and $N_p = 1,000$), the AOP solution matches the best solution amongst the MINLP solvers on up to 99% of the loans. Also for moderately large N, BARON's solution is often quite inaccurate, meaning it is significantly suboptimal compared to both BONMIN and the AOP. This highlights the well-known fact that MINLP solvers' success can vary in practice depending upon the particular problem. A priori, the computational performance of a particular MINLP solver on a problem is difficult to anticipate. In terms of computational time, the AOP generally takes several magnitudes less time than the MINLP solvers for moderately sized problems. In some cases, the AOP is as much as $5 - 6$ orders of magnitude faster than the MINLP solvers. BONMIN often foregoes branch and bound for larger problems, opting for its diving heuristic. In contrast to BONMIN's branch and bound method and the AOP, there is no theoretical guarantee on the accuracy of the diving heuristic for our class of problems. The diving heuristic is a type of relaxation method which solves a sequence of continuous optimization problems and rounds them to find an integer solution. Finally, for larger problems (e.g., selecting $N = 2,500$ out of $N_p = 10,000$ available loans or selecting $N = 25,000$ out of $N_p = 100,000$ available loans), the problem becomes computationally intractable for the MINLP solvers due to the high-dimensionality. In contrast, the AOP can efficiently and accurately solve such large-scale problems.

4.1 One-period Model

We consider the problem of selecting a portfolio of N loans from a pool of N_p available loans. The loan feature space $\mathcal{Y} = [-1, 1] \subset \mathbb{R}$ and $X_0 \in \{-1, +1\}$ (i.e., a "good" and a "bad" economy). The loans only have default risk (no prepayment) and default with transition probability:

$$h_\theta(\mathrm{d}, \mathrm{o}, y, x) = \frac{\exp(-3 + y + x)}{1 + \exp(-3 + y + x)}.$$

The one-period model is a special case of our model class (1) with $T = 1$. If $x = 1$, the loss given default is 50%. If $x = -1$, the loss given default is 30%. If a loan does not default, its return is $c + \frac{y+1}{10}$ (riskier loans have higher interest rates). The N_p available loans are drawn uniformly on $\mathcal{Y}$. In this one-period model, the AOP's objective and gradient can be evaluated in closed-form (up until quadrature), even if, for instance, $\mathcal{Y}$ was multi-dimensional and X was continuous-valued. The integer program's objective and gradient can also be evaluated in closed-form for many cases such as mean-variance, exponential utility, characteristic function, and moments. However, for many other functions (such as log utility or a tranche payoff), the objective and gradient for the true optimization problem must be evaluated via Monte Carlo simulation. This is yet another advantage for the AOP over the integer program in the case of the one-period model.

To solve the AOP, one has to choose a grid. We choose a grid of 200 points. The grid points are chosen as the centroids from k-means clustering of the N_p available loans. Since the AOP is measure-valued, one may not end up with an integer solution at all grid points. Thus, given the AOP, one has to decide how to "round" the AOP to get an integer solution. We use the simple heuristic of taking the ceiling of the grid points with the largest values such that total portfolio size is N and setting the remainder to their floor. We consider several problems for various sizes N and N_p. We first examine the cases of a mean-variance portfolio, an optimal portfolio under exponential utility, and the maximization of the expected return of an equity tranche of an asset-backed security. Finally, we present an example where the goal is to construct a geographically diverse portfolio.

4.1.1 Mean Variance Portfolio

We compare the performance of the AOP versus the integer program solvers BONMIN and BARON for finding the mean-variance portfolio. First, we consider the problem of choosing 250 loans from an available pool of $1,000$ loans. We require the portfolio to have a minimum expected return of 4% and the objective is to minimize the variance. The comparison of the MINLP solvers and the AOP is given in Table 1. The table reports the time for the solver, the reason it stopped (either reached the maxtime, minimum stepsize, minimum objective value change, or ran out of the 16GB of available RAM), the objective function evaluated at the solution, and a comparison of each solution with the AOP solution. The comparison of a solution with the AOP solution is the number of loans (out of N_p total loans) on which the two solutions agree. All MINLP solver solutions are suboptimal compared to the AOP solution. This means that the true objective is (lower) when it is evaluated at the AOP solution rather than the MINLP solution. Furthermore, the AOP solution requires 1-2 seconds of computation time, while the integer program solvers take orders of magnitude longer. We run the MINLP solvers for various lengths of times to show how their solutions evolve over time. That is, we stop the MINLP solvers early (short of their convergence criterion) to see how accurate their solution is. The first run reported for an integer program solver is for the minimum time required for one iteration of the solver. BONMIN takes 10 minutes for its first iteration while BARON takes almost an hour in its first iteration. BONMIN is still suboptimal compared to the AOP solution even after several hours. BONMIN's solution agrees with the AOP solution for over 97% of the loans. This simple example highlights the computational challenges faced by even the leading global integer solvers, even at this relatively "small" portfolio scale.

Solver	Time (s)	Exitflag	True objective	Comparison with AOP
BONMIN	500	Max time	37.5674	97.0 %
BONMIN	1000	Max time	37.4351	97.2 %
BONMIN	3111	Max time	37.4351	97.2 %
BONMIN	10000	Max time	37.4351	97.2 %
BARON	3051	Max time	97.7838	67.4 %
AOP	1	Min stepsize	37.3758	100 %

Table 1: Performance comparison between MINLP solvers and AOP for the selection problem of $N = 250$ loans out of a pool of $N_p = 1,000$ loan for a mean-variance optimization problem under a one-period model.

Next, we increase the size of the problem by one order of magnitude, i.e., we select 2,500 loans out of 10,000. As expected, the performance gain obtained through the AOP is even more pronounced for this larger problem. Out of the four integer programming cases we ran (three runs of BONMIN for different maximum time intervals and a single run of BARON), only two produced an output (the others failed to find a solution before the maximum allowed time), and both outputs were suboptimal compared to the AOP. In all cases, the MINLP solvers were significantly slower than the AOP. Furthermore, BARON eventually ran out of RAM and failed to produce a solution. The results are displayed in Table 2.

Solver	Time (s)	Exitflag	True objective	Comparison with AOP
BONMIN	500	Max time	No output	0 %
BONMIN	1000	Max time	3444.82	99.62 %
BONMIN	10000	Max time	3444.82	99.62 %
BARON	$>$10000	Out of RAM	No output	0 %
AOP	2	Min stepsize	3435.88	100 %

Table 2: Performance comparison between MINLP solvers and AOP for the mean-variance selection problem of $N = 2,500$ loans out of a pool of $N_p = 10,000$ loans under a one-period model.

4.1.2 Optimal Portfolio under Exponential Utility

Next, we conduct similar tests for a portfolio which maximizes the exponential utility $1 - \mathbb{E}[\exp(\gamma R_P^N)]$ where $\gamma = -1$. Table 3 presents the solver results for the problem of choosing $N = 250$ loans from a pool of $N_p = 1,000$ loans. BONMIN and AOP agree on 99.8% of the loans, with the AOP having a slightly better solution. BARON's solution does not perform as well in this case. BONMIN is roughly 90 times slower than AOP, with BARON being $1,000$ times slower. Table 4 contains results for the problem of choosing $N = 2,500$ loans from a pool of $N_p = 10,000$ loans. BONMIN is more than 4 orders of magnitude slower than the AOP and its solution is suboptimal compared to the AOP's solution. For even larger N_p, the computational time for BONMIN increases at a dramatic rate. Table 5 considers the case of selecting $N = 5,000$ loans out of a pool of $N = 25,000$ loans. BONMIN takes over 5 days to finish it, and agrees well with the AOP solution, although it is again slightly suboptimal compared to the AOP. Due to BARON's poor performance even for $N_p = 1,000$ and the long lengths of time, we did not run BARON for the later tests with larger portfolio sizes.

4.1.3 Optimal Portfolio backing an Asset-backed Security

We now consider the problem of choosing a portfolio to maximize the expected return of an equity tranche of an asset-backed security. The payoff to the equity tranche holder is simply a call option with strike

Solver	Time (s)	Exitflag	True objective	Comparison with AOP
BONMIN	93	Search completed	.6628567	99.8 %
BARON	1,020	Search completed	.658	76.0 %
AOP	1	Min stepsize	.6628569	100 %

Table 3: Performance comparison between MINLP solvers and AOP for the selection problem of $N = 250$ loans out of a pool of $N_p = 1,000$ loans for a portfolio which maximizes exponential utility.

Solver	Time (s)	Exitflag	True objective	Comparison with AOP
BONMIN	33,209	Max time	.663122	97.5 %
AOP	1	Min stepsize	.663147	100 %

Table 4: Performance comparison between BONMIN and AOP for the selection problem of $N = 2,500$ loans out of a pool of $N_p = 10,000$ loans for a portfolio which maximizes exponential utility.

Solver	Time (s)	Exitflag	True objective	Comparison with AOP
BONMIN	455,997	Max time	.6634706	99.9 %
AOP	1	Min stepsize	.6634711	100 %

Table 5: Performance comparison between BONMIN and AOP for the selection problem of $N = 5,000$ loans out of a pool of $N_p = 25,000$ loans for a portfolio which maximizes exponential utility.

K. For the numerical examples in this section, the value $K = 4\%$ is used. Tables 6, 7, and 8 compare the performance of the AOP with BONMIN for various sizes of N. In all cases, BONMIN's solution is suboptimal compared to the AOP, and many orders of magnitude slower.

Solver	Time (s)	Exitflag	True objective	Comparison with AOP
BONMIN	1218	Max time	.0592	99.4 %
AOP	5	Min stepsize	.0594	100 %

Table 6: Performance comparison between BONMIN and AOP for the selection problem of $N = 125$ loans out of a pool of $N_p = 500$ loans for a portfolio maximizing expected return of equity tranche of an MBS.

Solver	Time (s)	Exitflag	True objective	Comparison with AOP
BONMIN	2100	Max time	.05949	99.4 %
AOP	5	Min stepsize	.05952	100 %

Table 7: Performance comparison between BONMIN and AOP for the selection problem of $N = 250$ loans out of a pool of $N_p = 1,000$ loans for a portfolio maximizing expected return of equity tranche of an MBS.

Solver	Time (s)	Exitflag	True objective	Comparison with AOP
BONMIN	19440	Max time	.05856	99.4 %
AOP	5	Min stepsize	.05857	100 %

Table 8: Performance comparison between BONMIN and AOP for the selection problem of $N = 2,500$ loans out of a pool of $N_p = 10,000$ loans for a portfolio maximizing expected return of equity tranche of an MBS.

4.1.4 Geographic Diversification of a Loan Portfolio

Economic conditions at two geographic locations tend to be less correlated the more distant the geographic locations are. For instance, a mortgage in California is less correlated with a mortgage in Florida than with another mortgage in California. We consider the problem of forming a loan portfolio diversified across geographic locations according to a mean-variance criterion. 401 geographic locations (roughly the number of Metropolitan Statistical Areas) are included and the common factor X is a taken to be a 401-dimensional Gaussian random variable with covariance $\frac{1}{2} + \frac{1}{2}\exp(-\frac{1}{1000} \times$ distance between location i and location j$)$. At each location, 24 loans are available at interest rates of $3.125, 3.375, \ldots, 8.875$ percent. Figure 1 shows the geographic locations for the loans. The one-period model admits a closed-form distribution, which is important since accurately simulating a high-dimensional covariance structure would require a large number of Monte Carlo samples. Fitting a high-dimensional covariance matrix to data is prone to significant overfitting; therefore, we have introduced the reduced-form covariance which depends upon the distance between two locations.

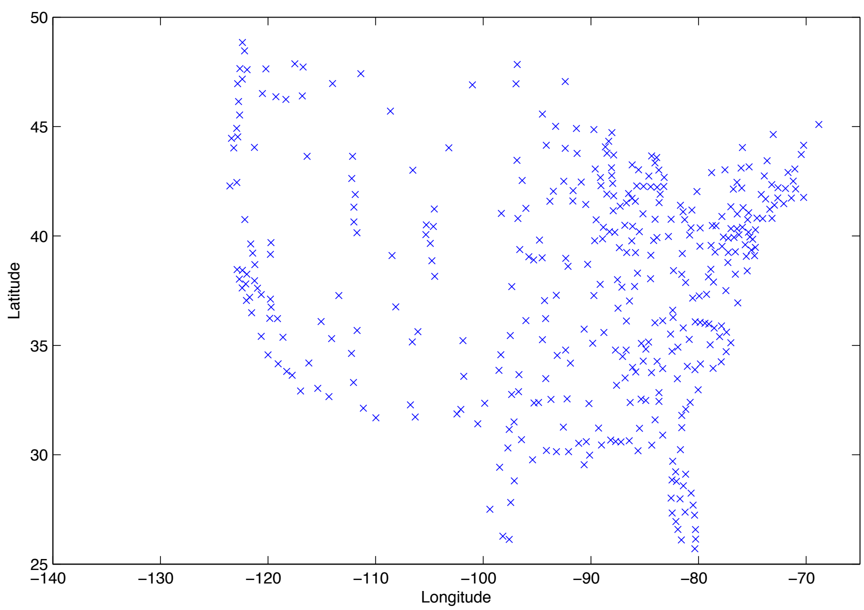

Figure 1: Geographic locations of the loans. The correlation between loans decreases with distance.

Let $y = (y_1, y_2)$ where y_1 is the interest rate and y_2 is the geographic location. The default probability for each loan is:

$$h_\theta(\mathrm{d}, \mathrm{o}, y, x) = \frac{\exp(g(y_1) + \sum_{i=1}^{401} x_i \mathbf{1}_{y_2=i})}{1 + \exp(g(y_1) + \sum_{i=1}^{401} x_i \mathbf{1}_{y_2=i})},$$

where g is a third-order polynomial with coefficients $-4.221, -.4108, .2120$, and $-.0149$ (which were fitted to the historical subprime mortgage default data described in Section 4.2.2 below). In total, there are $N_p = 9624$ available loans to select from and the goal is to select a minimum variance portfolio of $N = 2,500$ loans. Due to there being many geographic locations, both the AOP and the integer program are high-dimensional. As mentioned earlier, even high-dimensional continuous optimization problems can be extremely computational challenging. In order to deal with this, we implement block coordinate descent. Block coordinate descent divides the solution into subsets and cycles through these subsets. At each iteration, it fixes the solution on all but one subset. It solves the problem for that much smaller subset of "free variables". Block coordinate descent has been proven to converge for continuous convex optimization problems (see Tseng (2001)). Table 9 displays the results. BONMIN's solution agrees with the AOP on 99.5% of the loans. The AOP's solution is slightly less optimal compared to BONMIN's solution. This example could be extended by applying block coordinate descent to the AOP in order trace out the entire efficient frontier for the mean-variance portfolio.

Solver	Time (s)	Exitflag	True objective	Comparison with AOP
BONMIN	894	Max time	.001409	99.5 %
AOP	12	Max iterations	.001410	100 %

Table 9: Performance comparison between BONMIN and AOP for the selection problem of $N = 2,500$ loans out of a pool of $N_p = 9,624$ loans for a minimum variance portfolio.

4.2 Multi-period Model

In this section, we optimize portfolios of mortgage loans under the full dynamic model (1), which will be fitted to actual mortgage default and prepayment data. The pool of mortgages available for portfolio selection is also drawn from actual mortgage data. All interest payments and coupon payments are assumed to be monthly, and the times $t = 1, \ldots, T$, are months. The risk-free rate is chosen to be a constant $r_t = .01$. The common factor X that drives correlation in the mortgage portfolio includes the national mortgage rate and the national unemployment rate. Each element $i \in \{u, m\}$ of X is chosen to be an independent, discrete-time CIR process:

$$\begin{aligned}
\hat{X}_t^i &= \hat{X}_{t-1}^i + \kappa^i (m^i - \hat{X}_{t-1}^i) + \sigma^i \sqrt{\hat{X}_{t-1}^i} \epsilon_t^i, \\
X_t^i &= \max(\hat{X}_t^i, 0), \qquad (12)
\end{aligned}$$

where $\{\epsilon_t^i\}_{t\in I, i=\mathrm{u},\mathrm{m}}$ are i.i.d. standard normal variables. The value of X_t^i is not allowed to go negative since this would result in the square root of a negative number. The subscripts u and m refer to the unemployment rate and mortgage rate, respectively. The method of maximum likelihood is used to fit the unemployment rate to monthly national unemployment data for the United States from 1948 to 2014.[6] The mortgage rate is fitted to monthly 30-year fixed national mortgage rate data for the United States from 1971 to 2014. We note that many other model choices could be made for the stochastic process X; our framework imposes no assumptions on the form of X. Another possible choice for X is a vector autoregression.

4.2.1 Mean-Variance portfolio

We consider the problem of selecting a portfolio of N agency mortgages from a pool of N_p available agency mortgages according to a mean-variance criterion. Specifically, the problem is to minimize the variance of the portfolio return subject to the expected portfolio return being greater than 1%.

Since agency mortgages have very low default rates (and, in fact, Fannie Mae and Freddie Mac will compensate the MBS investor for any loss from default), we only model prepayments. Using the method of maximum likelihood, the parameter θ for h_θ is fitted to a data set of Freddie Mac mortgages over the time period $1999-2014$ which consists of 16 million mortgages. The mortgages are all 30-year fixed rate and fully amortizing. The loan-level feature space $\mathcal{Y}$ includes the FICO score, first-time homebuyer flag, number of units, occupancy status, combined loan-to-value ratio, loan-to-value ratio, initial interest, prepayment penalty flag, property type, loan purpose, number of borrowers, debt-to-income ratio, and geographic location (Metropolitan Statistical Area). We take

$$h_\theta(\mathrm{p},\mathrm{o},y,x) = \frac{1}{1+\exp(\theta_0+\theta_Y\cdot y+\theta_X^{\mathrm{u}}\cdot x^{\mathrm{u}}+\theta_X^{\mathrm{m}}\max(y^i-x^m,0))}, \tag{13}$$

where y^i is the interest rate for the mortgage. A dimension reduction for the limiting law $\bar{\mu}^N$ can be performed via the coordinate transformation $w=(z,y^i)=(\theta_Y\cdot y,y^i)\in\mathbb{R}^{d_W}$, which greatly reduces the computational expense of simulating $\bar{\mu}^N$ as well as finding the optimal portfolio. We refer to this coordinate transformation as the "low-dimensional transformation". In this case, $d_W=2$, much smaller than the original dimension of $\mathcal{Y}$. We will refer to the variable $z=\theta_Y\cdot y$ as the "prepayment inclination". Note that this transformation, which reduces the dimension of the problem, is an exact transformation (no accuracy is lost). The "prepayment inclination" is a variable (which is a linear combination of many features of the loan) influencing how likely that loan is to prepay based upon the characteristics of the loan (such as FICO, LTV ratio, etc.).

To solve for the AOP, one has to choose a grid. A natural grid for problems where N_p is small or moderately sized (hundreds or several thousand loans) is the actual coordinates in $\mathbb{R}^{d_W}$ of the available loans $w^1,\dots,w^{N_p}$. We use the same heuristic for rounding described earlier in the one-period model section. Finally, we note that the AOP for the mean-variance case with this choice of grid points bares close resem-

[6]All parameter fits are available from the authors upon request.

blance to the well-known heuristic of relaxing an integer program: that is, solving the continuous version of an integer program and rounding to yield a suitable integer solution. However, such relaxations are often heuristic, while in our case the AOP has a theoretical guarantee to converge. Moreover, for large-scale problems, the relaxation heuristic will suffer because even *continuous* nonlinear optimization problems become computationally challenging in high-dimensions. However, such large-scale problems can be solved via the AOP by choosing a grid whose number of mesh points is much smaller than N_p (see Section 4.3).

Solver	Time (s)	Exitflag	True objective	Comparison with AOP
BONMIN (diving heuristic)	200	Max time	0.089816	98.2 %
BONMIN (diving heuristic)	10,000	Max time	0.089816	98.2 %
BONMIN (branch & bound)	200	Max time	0.280683	76.8 %
BONMIN (branch & bound)	5,000	Max time	0.090541	89.2 %
BONMIN (branch & bound)	10,000	Max time	0.090518	91.6 %
BONMIN (branch & bound)	50,000	Max time	0.090447	91.8 %
BARON/CPLEX+PreProc.	3,600	Max time	0.524601	67.4 %
BARON/CPLEX+PreProc.	12,700	Max time	0.092638	80.8 %
BARON/CPLEX+PreProc.	52,700	Max time	0.091490	95.8 %
BARON/CPLEX	3,600	Max time	0.094147	84 %
BARON/CPLEX	12,700	Max time	0.092620	80.8 %
BARON/CPLEX	52,700	Max time	0.091183	95.8 %
BARON/CBC+PreProc.	12,700	Max time	0.10669	68.6 %
BARON/CBC+PreProc.	52,700	Max time	0.10669	68.6 %
BARON/CBC	12,700	Max time	N/A	0 %
BARON/CBC	52,700	Max time	0.10669	68.6 %
AOP	2	Min stepsize	0.089815	100 %

Table 10: Performance comparison between MINLP solvers and AOP for the mean-variance selection problem of $N = 250$ loans out of a pool of $N_p = 1,000$ loans.

The comparison of the MINLP solvers and the AOP is given in Table 10. All MINLP solver solutions are suboptimal compared to the AOP solution, i.e., the true objective evaluated at the AOP solution has a better value. We run the MINLP solvers for various lengths of times to show how their solutions evolve over time. That is, we stop the MINLP solvers early (short of their convergence criterion) to see how accurate their solution is. BONMIN's diving heuristic is able to find a good solution relatively quickly, although still two orders of magnitude slower than the AOP. However, BONMIN's diving heuristic is still suboptimal compared to the AOP solution after $10,000$ seconds. BONMIN's branch and bound is still far off from the solution even after $50,000$ seconds (almost 14 hours), as it seems stuck in a local optimum. BARON is slower; and its solution is still quite far off after $52,700$ seconds (over 14 hours).

Figure 2 compares the performance of the AOP and BONMIN. The pool of available loans are marked by $\times$, the loans chosen by the AOP are marked by $\square$, and the loans on which the AOP and integer programs disagree are marked by $\circ$.

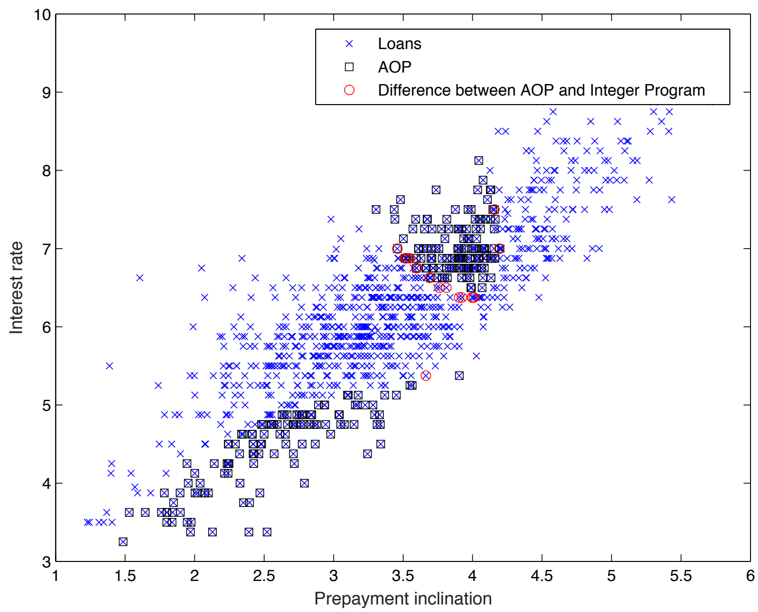

Figure 2: Comparison of AOP with BONMIN for mean-variance portfolio with $N = 250$ and $N_p = 1,000$.

4.2.2 Log-optimal portfolio

We now select a portfolio of N subprime mortgage loans from a pool of N_p available loans with the objective of maximizing the expected log utility of the portfolio return. Specifically, the objective is to maximize $\mathbb{E}[\log(\gamma R_P^N)]$ where R_P^N is the return of the portfolio P and γ is the risk aversion coefficient. For this example, $\gamma = 1$ (which is equivalent to maximizing the exponential growth in the "long run").

The subprime loans have both default and prepayment risk. We choose h_θ to be a multinomial logistic regression model. The transition functions $h_\theta(\mathrm{p}, \mathrm{o}, y, x)$ and $h_\theta(\mathrm{d}, \mathrm{o}, y, x)$ are given by

$$\frac{\exp(\theta_{0,p} + \theta_{Y,p} \cdot y + \theta^{\mathrm{u}}_{X,p} \cdot x^{\mathrm{u}} + \theta^{\mathrm{m}}_{X,p} \max(y^i - x^m, 0))}{1 + \exp(\theta_{0,d} + \theta_{Y,d} \cdot y + \theta^{\mathrm{u}}_{X,d} \cdot x^{\mathrm{u}}) + \exp(\theta_{0,p} + \theta_{Y,p} \cdot y + \theta^{\mathrm{u}}_{X,p} \cdot x^{\mathrm{u}} + \theta^{\mathrm{m}}_{X,p} \max(y^i - x^m, 0))},$$
$$\frac{\exp(\theta_{0,d} + \theta_{Y,d} \cdot y + \theta^{\mathrm{u}}_{X,d} \cdot x^{\mathrm{u}})}{1 + \exp(\theta_{0,d} + \theta_{Y,d} \cdot y + \theta^{\mathrm{u}}_{X,d} \cdot x^{\mathrm{u}}) + \exp(\theta_{0,p} + \theta_{Y,p} \cdot y + \theta^{\mathrm{u}}_{X,p} \cdot x^{\mathrm{u}} + \theta^{\mathrm{m}}_{X,p} \max(y^i - x^m, 0))},$$

respectively. The parameter θ is fitted using a data set containing over 10 million subprime mortgages. The data set was generously provided by the Trust Company of the West. The features include FICO credit score, LTV ratios, original balance, initial interest rate, loan type, default and prepayment times, and zip codes. Once the model has been fitted, we again use the low-dimensional transformation $w = (z, y^i) = (\theta_Y \cdot y, y^i) \in \mathbb{R}^{d_W}$, where z is now two-dimensional since we are considering both default and prepayment risk. In thise case, $d_W = 3$. The first dimension of w is the "default inclination", the second dimension is the "prepayment inclination", and the third dimension is the interest rate. The "default inclination" is a variable (which is a linear combination of many features of the loan) influencing how likely that loan is to default based upon the characteristics of the loan (such as FICO, LTV ratio, etc.). Similarly, the "prepayment inclination" is a variable (which is a linear combination of many features of the loan) influencing how likely that loan is to prepay.

N_p mortgages are drawn from the subprime mortgage data set for the portfolio problem. For convenience, we assume all mortgages are 15 year fixed-rate loans. When a default occurs, we assume that the loss given default is $\frac{1}{2}$. This means that the investor receives $\frac{1}{2}$ of the mortgage's balance at the time of default. The objective function (expected log utility) is evaluated via Monte Carlo simulation. We use $10,000$ Monte Carlo samples. For consistency, the same paths for the common factors X are used for both the true optimization problem as well as the AOP.

Table 11 compares the performance between BONMIN and the AOP. The two solutions agree on 98.8% of the loans and the AOP is two orders of magnitude faster. Figure 3 shows a comparison of the integer program and AOP solutions. The pool of available loans are marked by $\times$, the loans chosen by the AOP are marked by $\square$, and the loans on which the AOP and integer programs disagree are marked by $\circ$. Only two dimensions of the space $\mathbb{R}^{d_W}$ are displayed in Figure 3; the other dimension, "prepayment inclination", is not included in the figure. However, all three dimensions are used in the model and computational study.

Solver	Time (s)	Exitflag	True Objective	Nodes	Comparison with AOP
BONMIN	2419	Max time	-5.685347	NA	98.6 %
AOP	31	Min stepsize	-5.685306	NA	100 %

Table 11: Integer program and AOP performances where we choose $N = 250$ out of a pool of $N_p = 1000$ loans for a log-optimal portfolio.

4.3 Large-scale Optimization

In this section, we show how the AOP can tractably handle large portfolio optimization problems. By "large", we mean problems in the tens of thousands, hundreds of thousands, or even hundreds of thousands. To demonstrate the challenge of such large-scale problems, consider the problem of selecting N loans from an available universe of N_p loans. For the vast majority of problems, the objective and constraint functions in the true optimization problem (3) must be stochastically approximated via Monte Carlo simulation. Suppose that one performs L Monte Carlo simulations and that the objective and constraint functions are not path dependent (for instance, only depend upon the sum of discounted cashflows over all the times $t = 1, \ldots, T$). Then, one must store in memory and perform matrix operations on an $N_p \times L$ matrix. For large N_p and L, this matrix will not be storeable in memory of a typical desktop computer. For instance, if $N_p = 100,000$

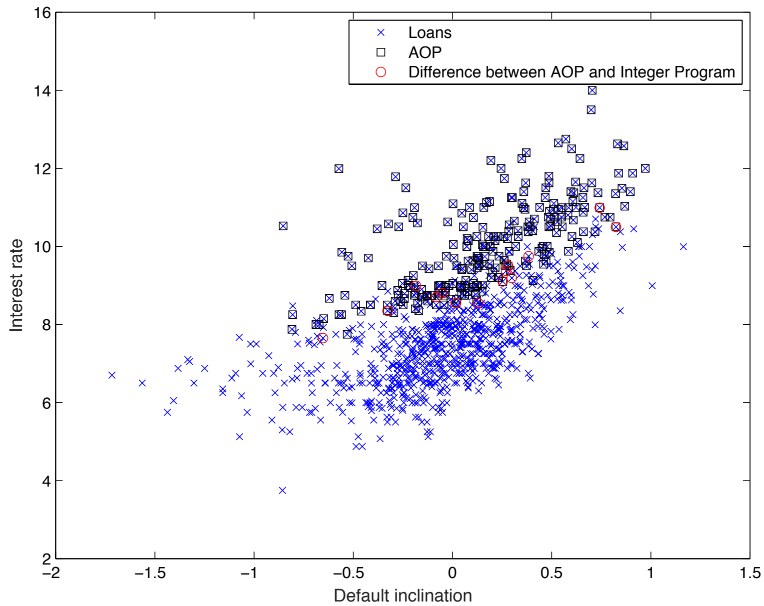

Figure 3: Comparison of AOP with BONMIN for log-optimal portfolio with $N = 250$ and $N_p = 1,000$.

and $L = 25,000$, the matrix would have 2.5 billion elements. Furthermore, matrix operations on such large matrices are extremely computationally expensive. These operations must be performed at each iteration of the optimization routine; on top of these computational challenges, one has an integer program of dimension N_p. Nonlinear integer programs with tens of thousands or hundreds of thousands of variables are generally considered intractable. Even with a cluster of computers, problems of this size would be computationally difficult and even intractable.

If the objective and constraint functions depend upon the path of the cashflows, then the problem becomes even more challenging. A matrix of size $N_p \times T \times L$ must be stored. If $N_p = 100,000$, $L = 25,000$, and $T = 360$, the matrix would have around 900 billion elements. A matrix of this size cannot be stored in memory, and a cluster would be needed to store the matrix as well as perform such large matrix operations. Furthermore, simply simulating the cashflow matrix is expensive.

A natural question might be whether the AOP does not suffer from some of the same challenges. In the last section, we used the actual data points for the available loans as the grid on which to solve the AOP. In that case, there would be N_p grid points. Even though the problem is a continuous optimization problem instead of a integer program, the matrix operations required for second-order continuous optimization methods (e.g., interior point methods) are still extremely expensive. For instance, the continuous optimization problem can be solved using first-order methods (e.g., gradient descent) on a cluster even if N_p is in the hundreds of thousands. However, first-order methods such as gradient descent have slow convergence rates. For large N_p, the AOP can actually be solved on a sparse grid, requiring only a few (hundreds or a few thousand) grid points. Using such a sparse grid, the AOP can be accurately solved in a matter of seconds on a single computer using second-order optimization methods. This is one of the key advantages of the AOP. Another option, which is available for certain problems and is implemented in Section 4.1.4, is block coordinate descent. However, whenever possible, reduction of the number of grid points through sparse grids will substantially decrease computational cost.

We propose three methods for choosing a sparse grid for the AOP. One is to simply perform k-means clustering on the loan features $Y^1, \ldots, Y^N$ and use the centroids of the clusters as the grid points. The fraction of the available pool at the k-th grid point is the fraction of the pool in the k-th cluster. A slightly more accurate algorithm is "repeated clustering", which is described in Appendix C and is also based upon k-means clustering. One disadvantage of using k-means clustering on the loan features $Y^1, \ldots, Y^{N_p}$ is that if $\mathcal{Y}$ multi-dimensional, certain dimensions may be more important than others and clustering using y will not ignore this. An alternative is to instead perform k-means clustering on the law of large numbers evaluated at each available loan: $(\bar{\mu}_t^1(u, Y^1), \ldots, \bar{\mu}_t^M(u, Y^1)), \ldots, (\bar{\mu}_t^1(u, Y^{N_p}), \ldots, \bar{\mu}_t^M(u, Y^{N_p}))$. $\bar{\mu}_t^m(u, Y^n)$ is the law of large numbers for the m-th Monte Carlo trial and the n-th available loan. M Monte Carlo trials are used.

In Table 12, we study the performance of the sparse grids for the AOP when selecting $2,500$ out of $10,000$ loans for a log-optimal portfolio. The fitted model of Section 4.2.2 is used. We only use the LLN for the AOP. The sparse grid with clustering on the law of large numbers uses $M = 500$ Monte Carlo trials. The actual optimization uses $L = 10,000$ Monte Carlo trials. We compare the solution using a sparse grid with the exact solution (i.e., $10,000$ grid points). We compare the sparse grid solutions with the exact solution in terms of the percent of loans on which they agree. The sparse grid is highly accurate even with a few hundred grid points.

Number of Clusters	Repeated clustering using Y^n Comparison with exact solution	Clustering using $\bar{\mu}$ Comparison with exact solution
100	95.6 %	98.3 %
150	96.5 %	98.5 %
225	97.2 %	99.2 %
300	97.6 %	99.0 %
525	98.2 %	99.4 %
750	98.4 %	99.6 %
900	98.6 %	99.5 %
1200	98.9 %	99.5 %
1350	99.0 %	99.6 %
10,000	100 %	100 %

Table 12: Comparison of solution using sparse grid with exact solution.

5 Conclusion

In this paper, we address large-scale optimization of a portfolio of loans. Such problems are computationally challenging since they involve high-dimensional nonlinear integer optimization. We approximate the optimization problem for a broad class of dynamic models of loan risk using weak convergence results. The solution to the approximate optimization problem is an *asymptotically optimal portfolio* (AOP) which we prove weakly converges to the solution of the true integer program as the size of the portfolio grows large. In a series of numerical studies using actual loan data, we compare the AOP against leading nonlinear integer program solvers. The AOP is highly accurate for large and even moderately-sized portfolios and is often several orders of magnitude faster than the best-in-class integer program solvers tested. Our method allows for tractable, large-scale data-driven optimization of loan portfolios and could be applicable to security selection problems in other asset classes.

A Proof of Theorem 3.1

The proof builds upon the weak convergence results for the class of models (1). A feasible sequence of portfolios P^N which weakly converges to the asymptotically optimal portfolio $\bar{P}^{\infty,*}$ is constructed. We show that if the optimal portfolio $P^{N,*}$ does not weakly converge to the asymptotically optimal portfolio $\bar{P}^{\infty,*}$, then $V^N(P^N) < V^N(P^{N,*})$ for some N. However, this is a contradiction, so the optimal portfolio $P^{N,*}$ must weakly converge to $\bar{P}^{\infty,*}$. Weak convergence is equivalent to convergence under the Prokhorov metric ρ. Thus, Theorem 3.1 means that $\rho(P^{N,*}, \bar{P}^{\infty,*}) \to 0$ and, by the triangle inequality, $\rho(P^{N,*}, \bar{P}^{N,*}) \to 0$. (The exact same proof approach can be used to show that $\rho(\bar{P}^{N,*}, \bar{P}^{\infty,*}) \to 0$. We omit the details for this in the proof since the steps are the same as used to show $\rho(P^{N,*}, \bar{P}^{\infty,*}) \to 0$.) The Prokhorov metric is the natural metric under which to study convergence for our problem due to the weak convergence results available for the model framework (1) and other similar models. Therefore, from a practical perspective, Theorem 3.1 implies that the asymptotically optimal portfolio $\bar{P}^{N,*}$ is accurate for large N.

We first prove the convergence of the finite optimal portfolio to the limiting optimal portfolio for the case without any constraints. Then, we extend the convergence result to the case with constraints. In this section, we will use the notation "$\Rightarrow$" to denote weak convergence.[7]

A.1 Convergence of Finite Optimal Portfolio to Limiting Optimal Portfolio without Constraints

Let (E, d) be a compact metric space. We will prove convergence of the finite optimal portfolios $P^{N,*}$ in the metric space $(\mathcal{M}(E), \rho)$ where ρ is the Prokhorov metric. This is the space of probability measures endowed with the topology of weak convergence. Let $E_d \subseteq E$ be a dense set in E (recall that a compact

[7] A probability measure $\nu^N \in \mathcal{M}(E)$ weakly converges to a probability measure $\nu \in \mathcal{M}(E)$ if and only if $\langle \phi, \nu^N \rangle_E \to \langle \phi, \nu \rangle_E$ for every continuous bounded function $\phi : E \to \mathbb{R}$.

metric space is separable). Define the spaces $\mathcal{M}^N(E_d) = \{\frac{1}{N}\delta_{a^1} + \cdots + \frac{1}{N}\delta_{a^N} : a^1, \dots, a^N \in E_d\}$ and $\mathcal{M}^N(E) = \{\frac{1}{N}\delta_{a^1} + \cdots + \frac{1}{N}\delta_{a^N} : a^1, \dots, a^N \in E\}$. Recall that the return of a portfolio P^N of size N is $R^N_{P^N}$ for $P^N \in \mathcal{M}^N(\mathcal{Y})$ and that if $P^N \Rightarrow P \in \mathcal{M}(\mathcal{Y})$, $R^N_{P^N} \xrightarrow{d} \bar{R}^\infty_P$ due to $\mu^N \xrightarrow{d} \bar{\mu}^P \in B^{T+1}$ where $B = \mathcal{M}(\mathcal{U} \times \mathcal{Y})$. Define the objective functions:

$$\begin{aligned} V^N(P^N) &= v_2\big(\mathbb{E}[v_1(R^N_{P^N})]\big), \\ V^\infty(P) &= v_2\big(\mathbb{E}[v_1(\bar{R}^\infty_P)]\big). \end{aligned} \tag{14}$$

We consider the optimization problems:

$$P^{N,*} = \arg\min_{P \in \mathcal{M}^N(\mathcal{Y})} V^N(P). \tag{15}$$

We wish to show that it converges to:

$$P^{\infty,*} = \arg\min_{P \in \mathcal{M}(\mathcal{Y})} V^\infty(P). \tag{16}$$

The optimization equation (15) is the true optimization problem where one can only hold loans in a portfolio in unit amounts, and there can only be N total number of loans held in the portfolio.

Assumption A.1. *Suppose that $\mathcal{Y}$ is compact, the functions h_θ, g, v_1, v_2, and q are continuous, and there exists a unique minimizer $P^{\infty,*}$ of (4) for $N = \infty$. Furthermore, suppose $f : B^{T+1} \times \mathbb{R}^{d_X} \to \mathbb{R}^{d_R}$ is of the form $f(\mu, x) = F(\mu, D(x))$ where $D : \mathbb{R}^{d_R} \to \mathbb{R}^{d_D}$ is a continuous bounded function and $F : B^{T+1} \times \mathbb{R}^{d_D} \to \mathbb{R}^{d_R}$ is continuous on $B^{T+1} \times \mathbb{R}^{d_D}$.*[8]

R^N_P is bounded since f is continuous, D is bounded, and $\mathcal{M}(\mathcal{U} \times \mathcal{Y})$ is compact. Since v_1 and v_2 are continuous and R^N_P is bounded, we also have that $V^N(P^N) = v_2\big(\mathbb{E}[v_1(R^N_{P^N})]\big) \to V^\infty(P^\infty) = v_2\big(\mathbb{E}[v_1(\bar{R}^\infty_{P^\infty})]\big)$ if $P^N \Rightarrow P^\infty$. Similarly, $J^N(P^N) \to J^\infty(P^\infty)$ if $P^N \Rightarrow P^\infty$.

Note that we do not make any restriction that $P^{N,*}$ must be unique. Also, from Lemmas B.5 and B.4, $V^N(P)$ and $V^\infty(P)$ are continuous on $\mathcal{M}^N(\mathcal{Y})$ and $\mathcal{M}(\mathcal{Y})$, respectively.

Lemma A.2. *The space $\mathcal{M}^N(\mathcal{Y})$ is compact.*

Proof. Since $\mathcal{M}(\mathcal{Y})$ is compact, $\mathcal{M}^N(\mathcal{Y}) \subset \mathcal{M}(\mathcal{Y})$ is also compact if $\mathcal{M}^N(\mathcal{Y})$ is closed. It is easy to see that $\mathcal{M}^N(\mathcal{Y})$ is closed for each N. Let $Z^N_m = \{\frac{1}{N}\delta_{a^1_m} + \cdots + \frac{1}{N}\delta_{a^N_m}\} \in \mathcal{M}^N(\mathcal{Y})$. Since $\mathcal{Y}$ is compact, $A_m = (a^1_m, \dots a^N_m)$ has a convergent subsequence $A_{m_k} \to A = (a^1, \dots, a^N) \in \mathcal{Y}$. If Z^N_m has a limit Z^N_∞, then $\frac{1}{N}\sum_{n=1}^N \phi(a^n_m) \to \langle \phi, Z^N_\infty \rangle$ as $m \to \infty$ for every continuous, bounded ϕ. Then, $\langle \phi, Z^N_\infty \rangle = \frac{1}{N}\sum_{n=1}^N \phi(a^n)$, which implies that $Z^N_\infty = \frac{1}{N}\delta_{a^1} + \cdots + \frac{1}{N}\delta_{a^N} \in \mathcal{M}^N(\mathcal{Y})$. ∎

Lemma A.3. *For any $P \in \mathcal{M}(\mathcal{Y})$, there exists a sequence $P^N \in \mathcal{M}^N(\mathcal{Y})$ such that $P^N \Rightarrow P$.*

[8] For the function $f(\mu, x) : B^{T+1} \times \mathbb{R}^{d_X} \to \mathbb{R}^{d_R}$, $\mu \in B^{T+1}$ and $x \in \mathbb{R}^{d_X}$.

Proof. Let $Y^1, \ldots, Y^N \in \mathcal{Y}$ be i.i.d. random variables with measure P. Define their empirical measure to be $\hat{P}^N = \frac{1}{N}\sum_{n=1}^{N} \delta_{Y^n} \in \mathcal{M}^N(\mathcal{Y})$. By Theorem 11.4.1 in Dudley (2002), $\hat{P}^N \Rightarrow P$ almost surely, i.e., $\mathbb{P}[\hat{P}^N \Rightarrow P] = 1$ where "$\Rightarrow$" denotes weak convergence. By contradiction, this implies that there exists a sequence $P^N \in \mathcal{M}^N(\mathcal{Y})$ such that $P^N \Rightarrow P$. Suppose no deterministic sequence $P^N \in \mathcal{M}^N(\mathcal{Y})$ existed such that $P^N \Rightarrow P$. Then, $\mathbb{P}[\hat{P}^N \Rightarrow P] = 1$ could not hold. Finally, since the measure P was arbitrary, we have the desired result. $\square$

Theorem A.4. *The sequence of optimal portfolios $P^{N,*}$ weakly converges to the limiting optimal portfolio $P^{\infty,*}$ as $N \to \infty$.*

Proof. Since $\mathcal{M}(\mathcal{Y})$ is compact and V^∞ is continuous on $\mathcal{M}(\mathcal{Y})$, a minimizer exists for the limiting problem. Furthermore, since we assumed in Assumption A.1 such a minimizer is unique, there exists a relatively open set $O_\nu \subset \mathcal{M}(\mathcal{Y})$ containing the unique minimizer such that all values of V^∞ outside of O_ν are uniformly separated from the value at the minimizer. To be precise, let $\nu > 0$ and $O_\nu = \{P \in \mathcal{M}(\mathcal{Y}) : \rho(P, P^{\infty,*}) < \nu\}$. Note that O_ν^c is compact (since a relatively closed subset of a closed set is also closed, and a closed subset of a compact set is compact) and $V^\infty(P^{\infty,*}) - V^\infty(P) < -\epsilon$ for $P \in O_\nu^c$ and some $\epsilon > 0$. By Lemma A.3, there exists a sequence of measures P^N which weakly converge to $P^{\infty,*}$. Moreover, by weak convergence, continuity of v, and boundedness of R_P^N, $V^N(P^N) \to V^\infty(P^{\infty,*})$ as $N \to \infty$.

The minimizers $P^{N,*} = \arg\min_{P \in \mathcal{M}^N(\mathcal{Y})} V^N(P)$ exist since $\mathcal{M}^N(\mathcal{Y})$ is compact and V^N is continuous on $\mathcal{M}^N(\mathcal{Y})$. Since $\mathcal{M}(\mathcal{Y})$ is compact, for every subsequence $P^{N_k,*}$, there exists at least one further subsequence $P^{N_{k_m},*}$ which weakly converges to some limit $P_2 \in \mathcal{M}(\mathcal{Y})$ as $m \to \infty$.[9] By weak convergence, we again have that $V^N(P^{N_{k_m},*}) \to V^\infty(P_2)$ as $m \to \infty$. Suppose that $P_2 \neq P^{\infty,*}$; this implies that $P_2 \notin O_\nu$ for some $\nu > 0$. Then, for some m_0, we have that for $m \geq m_0$:

$$\begin{aligned} V^{N_{k_m}}(P^{N_{k_m}}) - V^{N_{k_m}}(P^{N_{k_m},*}) &= [V^{N_{k_m}}(V^{N_{k_m}}) - V^\infty(V^{\infty,*})] + [V^\infty(P^{\infty,*}) - V^\infty(P_2)] \\ &+ [V^\infty(P_2) - V^{N_{k_m}}(P^{N_{k_m},*})] < \frac{\epsilon}{2} - \epsilon + \frac{\epsilon}{2} < 0. \end{aligned} \tag{17}$$

However, this is a contradiction since $P^{N_{k_m},*}$ is the minimizer of $V^{N_{k_m}}(P)$ for $P \in \mathcal{M}^{N_{k_m}}(\mathcal{Y})$, implying that it must be true that $V^{N_{k_m}}(P^{N_{k_m}}) - V^{N_{k_m}}(P^{N_{k_m},*}) \geq 0$. Therefore, $P_2 \in O_\nu$. Sending $\nu \to 0$, we have that, for every subsequence $P^{N_k,*}$ of $P^{N,*}$, there exists a further subsubsequence $P^{N_{k_m},*} \Rightarrow P^{\infty,*}$. Therefore, $P^{N,*} \Rightarrow P^{\infty,*}$. $\square$

[9] The sequence N_k is a subsequence of $\mathbb{N}$. The sequence N_{k_m} is a subsequence of N_k. The elements of the sequence N_k and N_{k_m} are respectively denoted as $N_1, N_2, \ldots$ and $N_{k_1}, N_{k_2}, \ldots$.

A.2 Proof for General Case with Constraints

Now, we extend the proof from the previous section to the case where there are constraints. The true optimization problem now becomes:

$$\begin{aligned} P^{N,*} \quad &= \quad \arg \min_{P \in \mathcal{M}^N(\mathcal{Y})} V^N(P), \\ &\text{s.t.} \qquad J^N(P) \leq c, \\ &\qquad\qquad q(P) \leq d, \end{aligned} \tag{18}$$

We wish to show that it converges to:

$$\begin{aligned} P^{\infty,*} \quad &= \quad \arg \min_{P \in \mathcal{M}^N(\mathcal{Y})} V^\infty(P), \\ &\text{s.t.} \qquad J^\infty(P) \leq c, \\ &\qquad\qquad q(P) \leq d, \end{aligned} \tag{19}$$

Assumption A.5. *In addition to the previous assumptions stated in Assumption A.1, assume that the optimization problems (18) and (19) are feasible.*

Let $F^N = \{P \in \mathcal{M}^N(\mathcal{Y}) : J^N(P) \leq c, q(P) \leq d\}$ and $F^\infty = \{P \in \mathcal{M}(\mathcal{Y}) : J^\infty(P) \leq c, q(P) \leq d\}$.

Lemma A.6. *The sets F^N and F^∞ are compact.*

Proof. The preimage of a continuous function on a closed set is also closed. In addition, the intersection of a compact set with a closed set is compact. Therefore, $F^N = \mathcal{M}^N(\mathcal{Y}) \cap \{J^N(P) \leq c\} \cap \{q(P) \leq d\}$ is compact. By the same reasoning, the F^∞ is also compact. □

Lemma A.7. *If P is in the interior of the feasible region F^∞, there exists a sequence P^N such that $P^N \Rightarrow P$ and P^N is feasible for the optimization problem (18) for every N.*

Proof. By Lemma A.3, there exists a sequence $P_0^N \in \mathcal{M}(\mathcal{Y})$ where $P_0^N \Rightarrow P$. As $N \to \infty$, $J^N(P_0^N) \to J^\infty(P) < c$ and $q(P_0^N) \to q(P) < d$. Then, there exists an N_0 such that $J^N(P_0^N) < c$ and $q(P_0^N) < d$ for $N \geq N_0$. Finally, let P^N be any sequence of feasible points for $N < N_0$ and set $P^N = P_0^N$ for $N \geq N_0$. □

Lemma A.8. *Any limit point P of a sequence of feasible portfolios P^N must be in the limiting feasible region F^∞.*

Proof. By compactness, we trivially have that $P \in \mathcal{M}(\mathcal{Y})$. Suppose that $P \notin F^\infty$. Then, one of the constraints will be violated. For instance, suppose the first constraint is violated, implying that $J^\infty(P) \geq c + \epsilon$ for some $\epsilon > 0$. However, since $J^N(P^N) \to J^\infty(P)$, this is a contradiction because it would imply that $J^N(P^N) > c$ for some N. The same reasoning can be applied to show the other constraints must be satisfied by P. □

Lemma A.9. *If $P^{\infty,*}$ is in the interior of the feasible region, $P^{N,*} \Rightarrow P^{\infty,*}$ as $N \to \infty$.*

Proof. The proof is exactly the same as in Theorem A.4. $\square$

Theorem A.10. *The optimal portfolios $P^{N,*}$ weakly converge to the limiting portfolio $P^{\infty,*}$ as $N \to \infty$.*

Proof. Due to Lemma A.9, all that remains is to prove the result when $P^{\infty,*}$ lies on the boundary of the feasible region. Construct the relatively open set $O_\nu \subset \mathcal{M}(\mathcal{Y})$ containing the unique minimizer such that all values of V^∞ outside of O_ν are uniformly separated from the value at the maximizer. To be precise, let $\nu > 0$ and $O_\nu = \{P \in F^\infty : \rho(P, P^{\infty,*}) < \nu\}$. Note that O_ν^c is compact (since a relatively closed subset of a closed set is also closed, and a closed subset of a compact set is compact) and $V^\infty(P) - V^\infty(P^{\infty,*}) > \epsilon$ for $P \in O_\nu^c$ and some $\epsilon > 0$. By continuity, one can choose a $\delta < \nu$ such that $V^\infty(P) - V^\infty(P^{\infty,**}) > \frac{\epsilon}{2}$ for any $P \in O_\nu^c$ and $P^{\infty,**} \in O_{2,\nu} \subset O_\nu$, where $O_{2,\nu} = \{P \in F^\infty : \rho(P, P^{\infty,*}) \leq \delta\}$. For the purposes of this proof, we will choose a point $P^{\infty,**} \in O_{2,\nu}$ which is in the interior of F^∞.

By Lemma A.7, there exists a sequence of measures $P^N \in F^N$ which weakly converge to $P^{\infty,**}$. Moreover, $V^N(P^N) \to V^\infty(P^{\infty,**})$ as $N \to \infty$.

Since $\mathcal{M}(\mathcal{Y})$ is compact and by Lemma A.8, for every subsequence $P^{N_k,*}$, there is at least one further subsubsequence $P^{N_{k_m},*}$ which weakly converges to some limit $P_2 \in F^\infty$ as $m \to \infty$. By weak convergence, we again have that $V^N(P^{N_{k_m},*}) \to V^\infty(P_2)$ as $m \to \infty$. Suppose that $P_2 \neq P^{\infty,*}$; this implies that $P_2 \notin O_\nu$ for some $\nu > 0$. Then, for some m_0, we have that for $m \geq m_0$:

$$\begin{aligned} V^{N_{k_m}}(P^{N_{k_m}}) - V^{N_{k_m}}(P^{N_{k_m},*}) &= [V^{N_{k_m}}(P^{N_{k_m}}) - V^\infty(P^{\infty,**})] + [V^\infty(P^{\infty,**}) - V^\infty(P_2)] \\ &+ [V^\infty(P_2) - V^{N_{k_m}}(P^{N_{k_m},*})] < \frac{\epsilon}{4} - \frac{\epsilon}{2} + \frac{\epsilon}{4} < 0. \end{aligned} \tag{20}$$

However, this is a contradiction since $P^{N_{k_m},*}$ is the minimizer of $V^{N_{k_m}}(P)$ for $P \in F^N$ and therefore $V^{N_{k_m}}(P^{N_{k_m}}) - V^{N_k}(P^{N_{k_m},*}) \geq 0$. Therefore, $P_2 \in O_\nu$. The result follows due to the same reasoning as in Theorem A.4. $\square$

B Limiting Laws

Assumption B.1. *Suppose that μ_0^N converges in distribution to $\bar{\mu}_0$, where $\bar{\mu}_0$ is deterministic, h is continuous, and $\mathcal{Y}$ is compact. Finally, also assume that $\sqrt{N}(\mu_0^N - \bar{\mu}_0)$ converges in distribution to $\bar{\Xi}_0$.*

Let $B = \mathcal{M}(\mathcal{U} \times \mathcal{Y})$. Provided Assumption B.1 holds, we have the following limiting laws for the system (1).

Theorem B.2. *The empirical measure μ^N converges in distribution to $\bar{\mu}$ in B^{T+1} as $N \longrightarrow \infty$, where $\bar{\mu}$ satisfies the equation:*

$$\bar{\mu}_t(u, dy) = \sum_{u' \in \mathcal{U}} h_\theta(u, u', y, X_{t-1}) \bar{\mu}_{t-1}(u', dy). \tag{21}$$

It is important to note that the law of large numbers is dynamic and is also a random equation; randomness enters through the factor X. The law of large numbers has a natural link with the original model (1). The function h_θ from (1) appears in the law of large numbers.

The law of large numbers can also be supplemented with a central limit theorem. Define the empirical fluctuation process $\Xi_t^N = \sqrt{N}(\mu_t^N - \bar{\mu}_t) \in W = \prod_{u=1}^{|\mathcal{U}|} S'(\mathbb{R}^{d_Y})$.[10] Like the law of large numbers, the central limit theorem is also dynamic. Randomness for the limiting process $\bar{\Xi}$ enters both through X and a martingale term M.

Theorem B.3. *Ξ^N converges in distribution to $\bar{\Xi}$ in W^{T+1} as $N \longrightarrow \infty$, where $\bar{\Xi}$ satisfies the equation:*

$$\bar{\Xi}_t(u, dy) \quad = \quad \sum_{u' \in \mathcal{U}} h_\theta(u, u', y, X_{t-1}) \bar{\Xi}_{t-1}(u', dy) + \bar{\mathcal{M}}_t(u, dy). \tag{22}$$

Given X, $\bar{\mathcal{M}}(u, dy)$ is a conditionally Gaussian process with zero mean and covariance:

$$\begin{aligned} Cov\Big[M_t(u_1, dy), M_t(u_2, dy)|X_{0:t-1}\Big] &= -\sum_{u' \in \mathcal{U}} h_\theta(u_1, u', y, X_{t-1}) h_\theta(u_2, u', y, X_{t-1}) \bar{\mu}_{t-1}(u', dy), \\ Var\Big[M_t(u, dy)|X_{0:t-1}\Big] &= \sum_{u' \in \mathcal{U}} h_\theta(u, u', y, X_{t-1})(1 - h_\theta(u, u', y, X_{t-1}) \bar{\mu}_{t-1}(u', dy), \end{aligned}$$

where $u_1 \neq u_2$.

Proofs for Theorems B.2 and B.3 can be found in Sirignano & Giesecke (2014).

Lemma B.4. *Suppose Assumption A.1 holds. Then, $V^\infty(P) = \mathbb{E}[v(\bar{R}_P^\infty)]$ is continuous on $\mathcal{M}(\mathcal{Y})$.*

Proof. It suffices to show that if $P_k \Rightarrow P$, $V^\infty(P_k) \to V^\infty(P)$. As earlier, let $\bar{\mu}^\nu$ be the law of large numbers with initial condition $\bar{\mu}_{t=0}(\text{o}, dy) = \nu$. Using the linearity of the law of large numbers (21), continuity of h, compactness of $\mathcal{Y}$, and induction, we have that for each X, $\bar{\mu}^{P_k} \Rightarrow \bar{\mu}^P$ as $P_k \Rightarrow P$. Then, for each X, $v(\bar{R}_{P_k}^\infty) \Rightarrow v(\bar{R}_P^\infty)$ as $P_k \Rightarrow P$. Note that $v(\bar{R}_P^\infty)$ is bounded since f, v are continuous, $\mathcal{M}(\mathcal{U} \times \mathcal{Y})$ is compact, and d is bounded. Using iterated expectations and the dominated convergence theorem, the result follows. $\square$

Lemma B.5. *Suppose Assumption A.1 holds. Then, $V^N(P) = \mathbb{E}[v(\bar{R}_P^\infty)]$ is continuous on $\mathcal{M}^N(\mathcal{Y})$.*

Proof. It suffices to show that if $P_k \Rightarrow P \in \mathcal{M}^N(\mathcal{Y})$, $V^N(P_k) \to V^\infty(P)$. To show this, it is enough to prove that for each X, $\mu_k^N \xrightarrow{d} \mu^N$ as $P_k \Rightarrow P$ where $\mu_{t=0}^N(\text{o}, dy) = P$. A convergence determining class of functions for B^{T+1} where $B = \mathcal{M}(\mathcal{U} \times \mathcal{Y})$ is $\zeta(\langle \phi_1, \nu \rangle, \ldots, \langle \phi_M, \nu \rangle)$, where $\zeta, \phi_1, \ldots, \phi_M$ are continuous and $\nu \in B^{T+1}$. (We have suppressed the notation $\langle \cdot, \cdot \rangle_{\mathcal{M}(\mathcal{U} \times \mathcal{Y})}$ for convenience.)

[10] S' is the space of tempered distributions.

$$\mathbb{E}[\zeta(\langle\phi_1,\mu_k^N\rangle,\ldots,\langle\phi_M,\mu_k^N\rangle)|X] = \mathbb{E}[\zeta(\frac{1}{N}\sum_{n=1}^{N}\sum_{t=1}^{T}\phi_1(t,Y_k^n,U_t^n),\ldots,\frac{1}{N}\sum_{n=1}^{N}\sum_{t=1}^{T}\phi_M(t,Y_k^n,U_t^n)|X]$$

$$= \sum_{u\in\mathcal{U}^{T\times N}}\prod_{n=1,\ldots,N,t\in I}\mathbb{P}[U_t^n=u_t^n|Y_k^n,X]\zeta(\frac{1}{N}\sum_{n=1}^{N}\sum_{t=1}^{T}\phi_1(t,Y_k^n,u_t^n),\ldots,\frac{1}{N}\sum_{n=1}^{N}\sum_{t=1}^{T}\phi_M(t,Y_k^n,u_t^n)). \quad (23)$$

Note that $\mathbb{P}[U_t^n=u_t^n|Y_k^n,X]$ is a continuous function of h_θ, which is itself a continuous function of Y_k^n. From Lemma A.2, P is of the form $\frac{1}{N}\delta_{a^1}+\cdots+\frac{1}{N}\delta_{a^N}$. Fix any $\epsilon>0$. Due to $P_k\Rightarrow P$, there exists a K such that for every $k\geq K$, $|\{y\in Y_k^1,\ldots,Y_k^N : \delta(y,a^{n'})\leq\epsilon\}| = |\{a\in a^1,\ldots,a^N : a=a^{n'}\}|$ for $n'=1,\ldots,N$.[11] Equation (23) is invariant under permutations of $(Y_k^1,\ldots,Y_k^N)$, and consequently has the limit:

$$\begin{aligned}
\lim_{k\to\infty} & \sum_{u\in\mathcal{U}^{T\times N}}\prod_{n=1,\ldots,N,t\in I}\mathbb{P}[U_t^n=u_t^n|Y_k^n,X]\zeta(\frac{1}{N}\sum_{n=1}^{N}\sum_{t=1}^{T}\phi_1(t,Y_k^n,u_t^n),\ldots,\frac{1}{N}\sum_{n=1}^{N}\sum_{t=1}^{T}\phi_M(t,Y_k^n,u_t^n)) \\
= & \sum_{u\in\mathcal{U}^{T\times N}}\prod_{n=1,\ldots,N,t\in I}\mathbb{P}[U_t^n=u_t^n|a^n,X]\zeta(\frac{1}{N}\sum_{n=1}^{N}\sum_{t=1}^{T}\phi_1(t,a^n,u_t^n),\ldots,\frac{1}{N}\sum_{n=1}^{N}\sum_{t=1}^{T}\phi_M(t,a^n,u_t^n)) \\
= & \mathbb{E}[\zeta(\langle\phi_1,\mu^N\rangle,\ldots,\langle\phi_M,\mu^N\rangle)|X].
\end{aligned}$$

This proves that for each X, $\mu_k^N \xrightarrow{d} \mu^N$ as $P_k\Rightarrow P$. Since f is continuous, $f(\mu_k^N,X)\xrightarrow{d} f(\mu^N,X)$. Since f is bounded, by dominated convergence theorem, it follows that $V^N(P_k)\to V^\infty(P)$ as $P_k\Rightarrow P$. □

C Repeated Clustering

After the low-dimensional transformation (described in Section 4), the available loans have characteristics $w\in\mathbb{R}^{d_W}$. The available loans for selection for the portfolio (which has size N) are $W=\{w^1,\ldots,w^{N_p}\}$.

- Normalize the points $W=w^1,\ldots,w^{N_p}$, initialize $i=0$, and set the stopping threshold τ. Initialize $\mathcal{R}$ as an empty list.
- For $i=0,\ldots,M$:
  (i) Find K clusters with centroids $\hat{c}_1,\ldots,\hat{c}_K$ from the points $\hat{W}=\{\hat{w}^1,\ldots,\hat{w}^{N_p}\}$ using k-means clustering. (Initialize the first centroid in the k-means clustering algorithm by randomly drawing a point from $\hat{w}^1,\ldots,\hat{w}^{N_p}$.) A loan w^n belongs to the cluster k^* if:

  $$k^* = \arg\min_{k=1,\ldots,K}||\hat{c}_k-\hat{w}^n||$$

  (ii) Let $C=(c_1,\ldots,c_K)$, where c_k is the unnormalized value for $\hat{c}_k$ (i.e., the reverse of the transformation used to normalize the data $w^1,\ldots,w^{N_p}$).

[11] This can be easily proven from the fact $P_k\Rightarrow P$ and using bump functions centered at each point $a^{n'}$.

(iii) Solve the AOP using the grid $\mathcal{R}^\frown C$. ("$a^\frown b$" is the concatenation of the lists a and b.) Round the solution at each grid point such that the solution is in increments of $\frac{1}{N}$ and sums to 1 across all grid points. Let the solution be p^i.

(iv) If $i \geq 1$, let Z be the subset of the loans $\hat{w}^1, \ldots, \hat{w}^{N_p}$ on which p^i and p^{i-1} disagree. Find K_i clusters from Z using k-means clustering and add the unnormalized centroids to the list $\mathcal{R}$.

(v) If $||p^i - p^{i-1}|| < \tau$, stop.

- Let the sparse grid be $\mathcal{R}^\frown C$.

The repeated clustering algorithm described above, although ad-hoc, is quite successful in practice. The k-means clustering finds grid points that minimize the within-cluster sum of squares and is a natural approach for finding a sparse grid that accurately captures the distribution of the available loans $w^1, \ldots, w^{N_p}$. However, in many cases, k-means clustering by itself is not enough since the optimization will place more importance on particular portions of the distribution. These portions of the space require a finer grid. Since a naive grid is not sufficient, two different k-means clusters from W will produce very different solutions for loans in these portions of the space. (Recall that k-means clustering is initialized randomly and only guarantees a local minimum; typically, two k-means clustering runs will produce different centroids.) Therefore, we place additional grid points (again using k-means clustering, but only on the loans that differ between the two consecutive solutions) in these portions of the space. This procedure is repeated until the solution is stable. The repeated clustering algorithm is therefore a nonuniform sparse grid which attempts to describe the distribution of W in a sparse manner and has a finer mesh where required by the optimization.

References

Akutsu, N., M. Kijima & K. Komoribayashi (2004), 'A portfolio optimization model for corporate bonds subject to credit risk', *Journal of Risk* **6**(2), 31–48.

Altman, E. (1996), Corporate bond and commercial loan portfolio analysis. Working Paper, Wharton Financial Institutions Center.

Andersson, F, H Mausser, D Rosen & S Uryasev (2001), 'Credit risk optimization with conditional value-at-risk criterion', *Mathematical Programming* **89**(B), 273–91.

Baesens, B. (2005), 'Neural network survival analysis for personal loan data', *Journal of the Operational Research Society* **56**(9), 1089–1098.

Banasik, J., J. Crook & L. Thomas (1999), 'Not if but when will borrowers default', *Journal of the Operational Research Society* **50**(12), 1185–1190.

Bastos, J. (2010), 'Forecasting bank loans loss-given-default', *Journal of Banking and Finance* **34**(10), 2510–2517.

Bellman, R. & S. Dreyfus (1959), 'Functional approximations and dynamic programming', *Mathematical Tables and Other Aids to Computation* pp. 247–251.

Bennett, P. (1984), 'Applying portfolio theory to global bank lending', *Journal of Banking and Finance* **8**, 153–169.

Bertsimas, D. & R. Shioda (2009), 'Algorithm for cardinality-constrained quadratic optimization', *Computational Optimization and Applications* **43**(1), 1–22.

Bienstock, D. (1996), 'Computational study of a family of mixed-integer quadratic programming problems', *Mathematical Programming* **74**(2), 121–140.

Blog, B, G Van der Hoeck, KA Rinnooy & GT Timmer (1983), 'The optimal selection of small portfolios', *Management Science* **29**(7), 792–798.

Bo, L. & A. Capponi (2014), Optimal investment in credit derivatives portfolio under contagion risk. *Mathematical Finance*, forthcoming.

Bokanowski, O., J. Garcke, M. Griebel & I. Klompmaker (2013), 'An adaptive sparse grid semi-lagrangian scheme for first order hamilton-jacobi bellman equations', *Journal of Scientific Computing* **55**(3), 575–605.

Bonami, P. & J. Goncalves (2012), 'Heuristics for convex mixed integer nonlinear programs', *Computational Optimization and Applications* **51**(2), 729–747.

Bonami, P, LT Biegler, AR Conn, G Cornuejols, IE Grossmann, CD Laird, J Lee, A Lodi, F Margot & A Waechter (2008), 'A polyhedral branch-and-cut approach to global optimization', *Discrete Optimization* **2**(5), 186–204.

Borzi, A. & V. Schulz (2009), 'Multigrid methods for pde optimization', *SIAM review* **51**(2), 361–395.

Burer, S. & A. Letchford (2012), 'Non-convex mixed-integer nonlinear programming: a survey', *Surveys in Operations Research and Management Science* **17**(2), 97–106.

Capozza, D., D. Kazarian & T. Thomson (1997), 'Mortgage default in local markets', *Real Estate Economics* **25**(4), 631–655.

Capponi, A & JE Figueroa-Lopez (2014), 'Dynamic portfolio optimization with a defaultable security and regime-switching', *Mathematical Finance* **24**(2), 207–249.

Capponi, A, JE Figueroa-Lopez & A Pascucci (2015), 'Dynamic credit investment in partially observed markets', *Finance and Stochastics* pp. 1–49.

Doya, K. (2000), 'Reinforcement learning in continuous time and space', *Neural computation* **12**(1), 219–245.

Dreyer, T., B. Maar & V. Schulz (2000), 'Multigrid optimization in applications', *Journal of Computational and Applied Mathematics* **120**(1), 67–84.

Dudley, R. (2002), *Real Analysis and Probability*, Vol. 74, Cambridge University Press.

Fabozzi, F., A. Bhattacharya & W. Berliner (2010), *Mortgage-backed securities: products, structuring, and analytical techniques*, John Wiley and Sons.

Gao, J. & D. Li (2013), 'Optimal cardinality constrained portfolio selection', *Operations Research* **61**(3), 745–761.

Giesecke, K, B Kim, J Kim & G Tsoukalas (2014), 'Optimal credit swap portfolios', *Management Science* **60**(9), 2291–2307.

Hemmecke, R., M. Koppe, J. Lee & R Weismantel (2010), *50 Years of Integer Programming 1958-2008*, Springer Berlin Heidelberg, chapter Nonlinear integer programming, pp. 561–618.

Khandani, A., A. Kim & A. Lo (2010), 'Consumer credit-risk models via machine-learning algorithms', *Journal of Banking and Finance* **34**(11), 2767–2787.

Kraft, H. & M. Steffensen (2008), 'How to invest optimally in corporate bonds: a reduced-form approach', *Journal of Economic Dynamics and Control* **32**, 348–385.

Kraft, H & M Steffensen (2009), 'Asset allocation with contagion and explicit bankruptcy procedures', *Journal of Mathematical Economics* **45**(1-2), 147–167.

Lewis, M. & S. Nash (2005), 'Model problems for the multigrid optimization of systems governed by differential equations', *SIAM Journal on Scientific Computing* **26**(6), 1811–1837.

Li, D. & X. Sun (2006), *Nonlinear Integer Programming*, Vol. 84 of *International Series in Operations Research and Management Science*, Springer.

Longstaff, F. & E. Schwartz (2001), 'Valuing american options by simulation: a simple least-square approach', *Review of Financial Studies* **14**(1), 113–147.

Meindl, P. & J. Primbs (2006), Corporate bond portfolio optimization with transaction costs. Working Paper, Stanford University.

Melennec, O (2000*a*), 'Asset backed securities: A practical guide for investors', *Societe Generale ABS Research* .

Melennec, O (2000*b*), 'CBO, CLO, CDO, a practical guide for investors', *Societe Generale ABS Research* pp. 1–14.

Mencia, J. (2012), 'Assessing the risk-return trade-off in loan portfolios', *Journal of Banking and Finance* **36**(6), 1665–1677.

Munos, R. & A. Moore (2002), 'Variable resolution discretization in optimal control', *Machine learning* **49**(2), 291–323.

Neumaier, A., O. Shcherbina, W. Huyer & T. Vinko (2005), 'A comparison of complete global optimization solvers', *Mathematical Programming* **103**(2), 335–356.

Paris, F. (2005), 'Selecting an optimal portfolio of consumer loans by applying the state preference approach', *European Journal of Operational Research* **163**(1), 230–241.

Saunders, D., C. Xiouros & S. Zenios (2007), 'Credit risk optimization using factor models', *Annals of Operations Research* **152**(1), 49–77.

Sirignano, J., A. Sadhwani & K. Giesecke (2015), Deep learning for mortgage risk. Working Paper, Stanford University.

Sirignano, J. & K. Giesecke (2014), Risk analysis for large pools of loans. Working Paper, Stanford University.

Stepanova, M. & L. Thomas (2002), 'Survival analysis methods for personal loan data', *Operations Research* **50**(2), 277–289.

Tawarmalani, M. & N. V. Sahinidis (2005), 'A polyhedral branch-and-cut approach to global optimization', *Mathematical Programming* **103**(2), 225–249.

Tseng, P. (2001), 'Convergence of a block coordinate descent method for nondifferentiable minimization', *Journal of Optimization Theory and Applications* **109**(3), 475–494.

Tsitsiklis, J. & B. Van Roy (2001), 'Regression methods for pricing complex american-style options', *IEEE Transactions on Neural Networks* **12**(4), 694–703.

van Handel, Ramon (2008), 'Hidden markov models', *Lecture Notes, Princeton University* .

Westgaard, S. & N. Van der Wijst (2001), 'Default probabilities in a corporate bank portfolio: a logistic model approach', *European Journal of Operational Research* **135**(2), 338–349.

Wise, M. B. & V. Bhansali (2002), 'Portfolio allocation to corporate bonds with correlated defaults', *Journal of Risk* **5**(1), 39–58.